

AD _____

GRANT NO: DAMD17-92-J-2006

TITLE: DNA Lesions in Medaka (O. Latipes): Development of a Micro-Method
for Tissue Analysis Using GC-MS

PRINCIPAL INVESTIGATOR: Donald C. Malins, Ph.D., D.Sc.

CONTRACTING ORGANIZATION: Pacific Northwest Research Foundation
Seattle, Washington 98122

REPORT DATE: August 15, 1995

TYPE OF REPORT: Final Report

19951211 084

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

DTIC QUALITY INSPECTED 1

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE August 15, 1995	3. REPORT TYPE AND DATES COVERED Final (30 Dec 91 - 29 Dec 94)	
4. TITLE AND SUBTITLE DNA Lesions in MEDKA (O. Latipes): Development of Micro-Method for Tissue Analysis Using GC-MS			5. FUNDING NUMBERS DAMD17-92-J-2006	
6. AUTHOR(S) Donald C. Malins				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Pacific Northwest Research Foundation Seattle, Washington 98122			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The application of gas chromatography-mass spectrometry with the complementary technique of Fourier-transform infrared spectroscopy to the analysis of DNA from cancer and noncancer tissue has provided a firm body of evidence demonstrating a pivotal role for free radicals in carcinogenesis. Substantial differences in radical-induced damage to DNA between cancer and noncancer tissue have been demonstrated in the female breast. Basically, the same type of damage was found in fish exposed to environmental contaminants, suggesting that the radical processes are phylogenetically conserved. The damage to DNA in the normal female breast has been shown to be a nonrandom progression culminating in a cancer-like phenotype in a substantial proportion of the women studied. Statistical evaluations of the DNA data derived from both analytical procedures allowed for the formulation of cancer probability models. Overall, the findings indicate that the genetic instability introduced into the DNA of normal tissues is pivotal in the etiology of cancer in diverse biological systems and that the DNA changes identified are promising biomarkers for cancer risk prediction.				
14. SUBJECT TERMS Fish, Molecular Biology, Environmental			15. NUMBER OF PAGES 68	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to *stay within the lines* to meet optical scanning requirements.

Block 1. Agency Use Only (Leave blank).

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract	PR - Project
G - Grant	TA - Task
PE - Program Element	WU - Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. (If known)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."

DOE - See authorities.

NASA - See Handbook NHB 2200.2.

NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.

DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.

NASA - Leave blank.

NTIS - Leave blank.

Block 13. Abstract. Include a brief (*Maximum 200 words*) factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code (*NTIS only*).

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

✓ Where copyrighted material is quoted, permission has been obtained to use such material.

NA Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

✓ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

NA In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

NA For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

NA In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

NA In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

NA In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Accession For	
NTIS CRACI	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

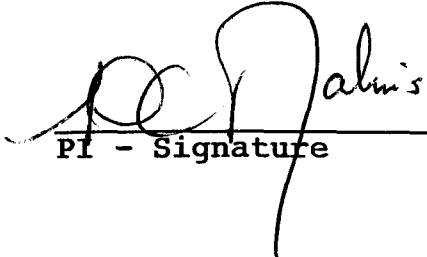
 8/14/95
PI - Signature Date

TABLE OF CONTENTS

I. Introduction	4
II. Body	7
A. Experimental Methods.....	7
1. Isolation of Medaka liver DNA.....	7
2. Preparation of trimethylsilyl derivatives	8
3. Synthesis of oxidized nucleotide bases	9
4. Gas chromatography-mass spectrometry/selected ion monitoring.....	9
5. Data collection and analysis	11
6. Isolation of female breast tissue DNA	11
7. Gas chromatography-mass spectral and Fourier transform-infrared spectral analysis	12
8. IR spectral and statistical analysis, and correlations with the Fapy-A/8-OH-Ade model	13
B. Results.....	18
1. Studies on Medaka Using Reduced Tissue Sample Weights	18
Table 1. <i>Improved DNA Extraction Methodology</i>	19
2. Application of the GC-MS/SIM Technique to the Analysis of Human Normal and Cancerous Breast Tissues	19
Figure 1. <i>Log₁₀ concentration of the base lesions (nmol/mg DNA) versus tissues analyzed</i>	22
Figure 2. <i>DNA base lesion values (mean ± standard error) for the cancerous (MNT and IDC) and normal female breast (RMT)</i>	23
Table 2. <i>Mean log₁₀ concentrations and log₁₀ ratios of concentrations of DNA-base lesions based on statistical models</i>	25
Figure 3. <i>The predicted probability of the cancerous origin of a tissues</i>	27

Figure 4. <i>A proposed scheme for the formation of the ring-opening (Fapy) derivatives and 8-OH-adducts in the female breast</i>	29
Table 3. <i>Classification of tissue sections using predictive model based on ratio of concentrations: (Fapy-A + Fapy-G)/(8-OH-Ade + 8-OH-Gua) or (Fapy-A)/(8-OH-Gua)</i>	32
3. Application of the FT-IR Technique to the Analysis of Human	
Normal and Cancerous Breast Tissues	35
Figure 5. <i>Mean normalized absorbance spectra of patients with cancer and without cancer and the statistical significance of cancer versus noncancer absorbances</i>	36
Table 4. <i>Comparison of spectral descriptive characteristics for cancer and noncancer patients</i>	37
a. Systematic shifts in the spectra toward a cancer-like phenotype	38
1) Figure 6. <i>Pearson correlation coefficient between normalized absorbance and distance-C from cancer library</i>	40
2) Figure 7. <i>The relationship of the distance of patients without cancer from the cancer library (distance-C) and normalized absorbances at frequencies 1172 and 1234 cm⁻¹</i>	41
b. Relationship of spectral features to base modifications	41
1) Table 5. <i>Pearson correlation of spectral descriptive characteristics with log₁₀ ratio of concentrations of Fapy-A to 8-OH-Ade</i>	42
c. Breast cancer risk model	42
1) Table 6. <i>Comparison of Factor Scores Between Patients with Cancer and Without Cancer (First Six Factors)</i>	43
2) Figure 8. <i>Predicted probability of cancer status using a logistic regression model</i>	44
3) Figure 9. <i>Groups defined by distance from cancer library</i>	45
d. Grouping of patients without cancer	45

Figure 10. <i>Mean spectra for patients with cancer and for groups of patients without cancer at three distances from the cancer library.....</i>	46
e. Results of factor analysis	48
1) Figure 11. <i>Relationship of cancer and noncancer groups based on factor analysis in the 1503-761-cm⁻¹ area</i>	48
2) Table 7. <i>Predictive Model for Cancer vs. Noncancer Status Based on a Logistic Regression Analysis of First Six Factors From Factor Analysis.....</i>	49
III. Conclusions	50
IV. Future Plans	58
V. References	59

I. INTRODUCTION

Carcinogenesis is considered to be a multistep process often that takes place as a consequence of exposure of cells to toxic, sublethal concentrations of chemicals capable of eliciting mutagenesis by virtue of changes in DNA base structures. Historically, DNA modifications arising from cytochrome P-450-catalyzed two-electron oxidation reactions, which yield bulky DNA base derivatives, have been regarded as significant in introducing mutagenic changes in DNA (1-3). However, such reactions also yield another product, H_2O_2 , which is readily converted to $\bullet OH$ in the Fe^{++} -mediated Fenton reaction (4, 5). The first evidence in a living system for the carcinogenic potential of radical-induced DNA oxidation came from studies of the marine flatfish English sole. Chronic, low level exposure to carcinogenic xenobiotics were capable of eliciting free radical oxidative DNA lesions in English sole inhabiting contaminated waterways (6). Such animals express a high incidence of liver cancer (up to about 30%) compared to animals from pristine environments (7). A striking increase in a range of $\bullet OH$ -induced base lesions was found which correlated both with exposure to xenobiotics and incidence of liver cancer compared to unexposed control animals (8). Subsequently, studies of $\bullet OH$ damage to DNA in a variety of human cancers and tissues have been conducted (9, 10). The results of these experiments agree closely with the English sole data and support the notion that $\bullet OH$ induced DNA damage is intimately associated with carcinogenesis.

We have also shown that substantial quantities of OH-induced base modifications are present in DNA from cancers of the female breast (9, 10) and that a characteristic pattern of such base modifications occurs in the normal and cancer-containing breast in relation to the redox

status of the tissue (9). Histologically normal (reduction mammoplasty) tissue was characterized by a high ratio of the ring-opening product 4,6-diamino-5-formamidopyrimidine (Fapyadenine; Fapy-A) to 8-hydroxy-adducts (e.g., 8-hydroxyguanine; 8-OH-Gua). A dramatic change in this relationship occurred favoring the substantial loss of ring opening products and a substantial increase in the 8-OH derivatives in the cancer-containing breast. Twenty-two statistical base models were formulated with a high sensitivity (e.g., 91%) and specificity (e.g., 97%) for classifying breast tissue as cancer or normal (e.g., 96% correct) (9). The dramatic shift in the proportions of the •OH-modified bases was attributed to changes in the redox potential of the breast that favor oxidative conditions and cancer formation (9). The statistical models most likely have predictive power for assessing future breast cancer risk. The radical-induced one-electron conversions of the DNA from the normal and cancer containing breast often produced one modification in several hundred normal bases. This is exceptionally high, notably relative to two-electron base modifications that form bulky adducts in the range of 1 in 10^7 to 1 in 10^9 normal bases, as shown by ^{32}P -postlabeling analysis of diverse tissues (11).

Djuric *et al.* (12) reported that the level of one of the base modifications previously described (5-hydroxymethyluracil; HMUra) (9) is elevated in peripheral blood of patients with breast cancer. Frenkel *et al.* (13) reported that 5-hydroxymethyl-2'-deoxyuridine elicited elevated circulating autoantibody levels reactive with this base structure in patients with breast cancer, those at high risk, and in those in whom breast cancer developed one or more years later. The circulating 5-hydroxymethyl-2'-deoxyuridine appeared to originate in breast tissue. These

findings support our original hypothesis regarding the potential use of the •OH-induced base modifications for assessing breast cancer risk (9, 10).

The importance of the 8-OH-adducts in carcinogenesis is well recognized: 8-OH-Gua has been implicated as a mutagen and carcinogen (14, 15), has an overwhelming effect in causing misreplication in template-directed DNA synthesis (15), and results in a loss of methylcytosines that participate in regulating the expression of genes (16). However, the ring-opening Fapy structures, as well as bulky adducts from two-electron oxidations, were shown to block DNA replication (17) and mRNA transcription (18).

Substantial oxidative modifications occurring in the DNA bases in breast carcinogenesis (9, 10) imply that significant modifications also occur in other aspects of the DNA, such as in the phosphodiester backbone and deoxyribose (4). Consequently, we studied DNA from the normal and cancer-containing breast using Fourier transform-infrared (FT-IR) spectroscopy in conjunction with the complementary technique of gas chromatography-mass spectrometry with selected ion monitoring (GC-MS/SIM). FT-IR spectroscopy is eminently suited to the identification of these structural changes (19). DNA from normal reduction mammoplasty tissue (RMT), invasive ductal carcinoma (IDC), and nearby microscopically normal tissue (MNT) was analyzed by FT-IR spectroscopy. Statistical models based on the spectral properties were obtained and compared to a statistical model previously used with the base modifications.

Substantial differences were found in the spectral properties of DNA from women with normal and cancerous breast tissue, indicating an ability to discriminate the cancerous tissue from noncancerous tissue with a sensitivity and specificity of 83%. More importantly, the

normal population was divided into subgroups in which a non-random progression was identified and a cancer-like DNA phenotype that was highly correlated ($r \geq 0.90$) with that of the patients with cancer was exhibited in 59% of the women. The spectral data, which highly correlated with the base-model data, were used to establish a model for predicting the probability of breast cancer. Consistent with the high cancer reoccurrence rate in the same breast, 8 of 10 of the MNTs remaining after tumor excision were classified as "cancer" using this model (10).

Progressive structural changes in the DNA of the normal female breast leading to a premalignant cancer-like phenotype in a high proportion of women are the basis for a new paradigm for understanding the etiology of breast cancer and predicting its occurrence at early stages of oncogenesis. Therapeutic strategies for potentially reversing the extent of DNA damage, which may be useful in disease prevention and treatment, are also suggested from the findings.

II. BODY

A. Experimental Methods

1. Isolation of Medaka liver DNA

During this grant period, the USABRD L provided samples of Medaka liver exposed to TCE together with controls. The samples were used for the isolation of pure DNA from as little as 2 to 5 mg of starting material. Our previous "macro" methods involved pretreatment of the sample with Proteinase K and RNase A, after which recovery of the nucleic acids from the cellular lysate was accomplished by employing phenol and chloroform to partition the nucleic

acids into an aqueous phase and the other cellular components, including proteins, into an organic phase.

Our laboratory employed a simpler and much more efficient "micro" method for isolation of the DNA (See Table 1). The Microprobe IsoQuick® Nucleic Acid Extraction Kit utilizes the properties of guanidine thiocyanate (GuSCN) to both disrupt the cellular integrity of the sample and, at the same time, inhibit the DNase and RNase activity. The GuSCN is then mixed with a non-corrosive reagent containing a nuclease-binding matrix. The aqueous and organic phases are separated by centrifugation and the DNA is precipitated with alcohol. Four milligrams of purified Chelex® 100 Resin (Bio Rad) is added to the DNA to remove any Fe^{++} that might be present. As a consequence, we have also found that the addition of Chelex® 100 Resin results in greater purification of DNA, yielding a higher A260/A280 spectral ratio. The DNA is quantitated in aqueous solution by its UV absorption at 260 nm using the relationship 1 absorbance unit = 50 $\mu\text{g}/\text{ml}$.

2. Preparation of trimethylsilyl derivatives

The procedure employed was a modification of that used previously (20). For example, samples of purified DNA are now made usually with 30 - 50 μg and done in either duplicate or triplicate, depending upon the availability of DNA. All are treated in evacuated sealed tubes at 140°C for 30 min. with 0.25 ml of concentrated formic acid (60%). The treatment with formic acid does not alter the structure of the nucleotide bases being studied. After hydrolysis, the samples are dried by lyophilization. The trimethylsilyl derivatives are produced in a 0.1 ml of

mixture of bis(trimethylsilyl)trifluor-acetamide (BSTFA) and acetonitrile (4:1) in polytetrafluorethylene-capped hypovials (Pierce Chemical Company) upon heating for 30 minutes at 140°C.

3. Synthesis of oxidized nucleotide bases

Several of the standards required for the GC/MS-SIM procedure were obtained from commercial sources; others had to be synthesized in our laboratories.

8-hydroxyadenine (8-OH-Ade) was synthesized, using 5-bromocytosine and 8-bromoadenine, respectively. These compounds were allowed to react with 95% concentrated formic acid at 140°C for 45 minutes. Excess unreacted formic acid was removed by nitrogen purge and the product was purified by recrystallization from water.

The 2,6-Diamino-4-hydroxy-5-formamidopyrimidine (Fapyguanine; Fapy-G) was synthesized from 2,5,6 triamino-4-hydroxypyrimidine sulfate and 80% formic acid at 60°C for one hour. Excess formic acid was removed with nitrogen. The product of the reaction was purified by recrystallization from water and purity established by GC-MS/SIM.

4. GC-MS/SIM

The analyses for oxidized nucleotide bases was conducted with a Hewlett-Packard Model 5890 microprocessor-controlled gas chromatograph interfaced to a Hewlett-Packard model 5970B Mass Selective Detector. The injector port and interface were both maintained at 260°C. The column was a fused silica capillary column (12.0 m, 0.2 mm inner diameter) coated with

cross-linked 5% phenylmethylsilicone gum phase (film thickness, 0.33 μm). The column temperature was programmed from 120° to 235°C at 10°C/min. after 2 min. at 120°C. Helium was used as the carrier gas with a linear velocity of 57.3 cm/s through the column. The amount of TMS hydrolysate injected onto the column was about 0.5 μg . Quantitation of TMS-nucleotide bases was done on the basis of the principal ion and confirmation of structure was undertaken using two qualifier ions.

Several major improvements have been made in the GC-MS/SIM methodology with the objective of "microtizing" the DNA analysis. A Hewlett-Packard Merlin Microseal® has been installed which decreases septum leakage and, at the same time, eliminates the presence of septum particulates in the injection liner which can cause activation of the liner. The MS detector has been changed to a Hewlett-Packard K-M® model, thereby increasing the sensitivity of the MS by about *five-fold*. The automatic injector has been altered so that the syringe pumps each sample a total of 12 times with a viscosity delay of 7 seconds. The BSTFA:ACN solvent is viscous enough to allow air bubbles to enter the syringe if a viscosity delay is not used. Though minute, these air bubbles can cause a 20-40% error in reproducibility.

A most important advance made in the GC-MS/SIM method is the automation of the quantitation procedure. The quantitation files for the base lesions have been integrated into one file, allowing each sample to be quantitated for all five lesions at once, rather than by 5 separate files. The results are individually checked for proper peak integration and then transferred to a MS Excel database in which the conversion from pg/ μl to nmol/mg DNA (including all recovery

and reproducibility factors) is automatically figured. The result is in tabular form, and the data is readily converted to a bar graph or other suitable depiction.

5. Data collection and analysis

Using the GC-MS/SIM methodology, characteristic ions (one principal ion and two qualifier ions) were employed to characterize the oxidized bases; however, as indicated, the principal ion was used for quantitation. All spectra were compared with spectra obtained from commercially obtained standards and authentic samples of TMS derivatives synthesized in our laboratories. The data obtained included SIM plots and derived mass spectra. On the basis of the GC-MS/SIM data, oxidized base concentrations in hepatic DNA were calculated and recorded as nmol/mg.

6. Isolation of female breast tissue DNA

Initially, RMT was obtained from 15 patients. The tissue from 10 patients was sequentially cut into 1 cm sagittal sections, 2 cm apart. The tissue from the remaining five RMT patients was divided into two sections. Two to 13 sections were obtained from each patient for a total of 70 samples. In addition, tumor (IDC) and nearby MNT were obtained from the cancerous breasts of 15 surgical patients. This group comprised 22 samples, 7 of which were matched pairs (IDC-MNT); the remainder were single biopsy specimens from either IDC tissue or MNT. The RMT from the patients who did not have cancer was also microscopically normal,

with the exception of occasional incidences of non-neoplastic changes (e.g., fibrocystic). The above-mentioned tissues were used for a study of the DNA bases via GC-MS/SIM.

Subsequently, normal RMT was obtained from 29 patients living in the Puget Sound region of Washington State; one to four sections were obtained from each patient, totaling 41 samples. Eighteen IDC tissues were obtained from the breasts of 18 surgical patients, and 11 nearby MNT specimens were obtained from an additional 10 surgical patients. The IDC and MNT specimens were obtained from local hospitals and The Cooperative Human Tissue Network (Cleveland, OH). The RMT and the MNT exhibited occasional incidences of non-neoplastic fibrocystic changes; other cellular changes were essentially absent. These tissues were used in studies involving FT-IR spectroscopy in conjunction with GC-MS/SIM.

Each of the excised tissues was frozen in liquid nitrogen and maintained at -80°C . The DNA was isolated from ~ 350 mg of tissue and the purity was established as previously described (9, 10). The DNA (yield: ~ 120 - $150\mu\text{g}$), dissolved in deionized water, was aliquoted into portions for GC-MS/SIM; $\sim 50\mu\text{g}$ (9, 10) and FT-IR spectroscopy ($\sim 20\mu\text{g}$) and the aliquots were dried completely by lyophilization, purged with pure nitrogen and stored in an evacuated, sealed glass vial.

7. Gas chromatographic-mass spectral and FT-IR analysis of breast tissues

Lyophilized DNA from female breast tissues was hydrolyzed, derivatized and analyzed by the same GC-MS/SIM method as used for Medaka liver (9, 10). Infrared (IR) spectra of 59 DNA samples were obtained with a Perkin-Elmer System 2000 Fourier Transform-IR

spectrometer (The Perkin-Elmer Corp., Norwalk, CT) equipped with an IR microscope and a wide range mercury-cadmium-telluride detector. The DNA was placed on a BaF₂ plate in an atmosphere with a relative humidity of less than ~60% and flattened to make a transparent film. Using the IR microscope, a uniform and transparent portion of the sample was selected to avoid a scattering or wedge effect in obtaining transmission spectra. Each analysis was performed in triplicate on 3-5µg of DNA and the spectra were computer averaged and two hundred and fifty-six scans at 4 cm⁻¹ resolution were performed to obtain spectra in a frequency range of 4000-700 cm⁻¹. Typically 3-5 minutes elapsed from when the glass vial was broken to when each spectrum was obtained. None of the spectra showed a 1703 cm⁻¹ band, which is indicative of specific base pairing. This fact indicates that the samples were free from water and had acquired a disordered form, the D-configuration (21-23).

The spectra were obtained in transmission units and converted to absorbance units for data processing. The Infrared Data Manager software package (The Perkin-Elmer Corp.) was used to control the spectrometer and to obtain the IR spectra. The GRAMS/2000 software package (Galactic Industries Corp., Salem, NH) was used to perform post-run spectrographic data analysis. Each spectrum was converted to a spreadsheet format giving a specific absorbance for every wavenumber between 4000 and 700 cm⁻¹.

8. IR spectral and statistical analysis, and correlations with the Fapy-A/8-OH-Ade model

The IR spectral and statistical analyses were used to test the hypotheses that systematic differences occur in DNA between normal and cancer tissue and that a progressive, non-random

modification of the DNA in the normal breast culminates in a cancer-like phenotype. This phenotype likely would constitute a premalignant biomarker for cancer development.

A baseline adjustment was made in all spectra to remove the effect of background reflectance. The mean across 11 wavenumbers, centered at the lowest point (2000 to 1700 cm^{-1} range), was subtracted from absorbances at each frequency.

The specimens varied in thickness, yielding diverse absorbances or spectral intensities unrelated to cancer or noncancer type. Because there was not a well established reference peak in the frequency range of interest to use for normalization, it was achieved by converting all absorbances to a constant mean intensity.

The region 1503 to 761 cm^{-1} -- a span of 743 wavenumbers -- was chosen as the primary region for analysis because it constituted the most common valley-to-valley absorbance range among all cancer and noncancer spectra. This area of the spectrum was chosen disregarding cancer/noncancer status. Absorbances at all wavenumbers were divided by the mean absorbance ranging from 1503 to 761 cm^{-1} for that spectrum. This resulted in a mean spectral intensity of 1.0 for each specimen. Unless otherwise stated, all analyses were carried out on these baselined and normalized spectra.

The cancer and noncancer spectral groups were compared by plotting means for each of the two groups at each frequency. A mean cancer and a mean noncancer DNA spectrum was obtained. For patients with multiple specimens, the normalized absorbances were averaged to yield a single mean spectrum per patient. At each spectral frequency, a *t*-test for the difference

between cancer and noncancer normalized absorbances was conducted. This yielded one P value per frequency. The method of Schweder and Spjøtvoll (24) was used to estimate the number of frequencies with cancer/noncancer differences that were unlikely to be random. Cancer and noncancer spectra were compared based on several other measures. Such a measure is distance-C, which was calculated as follows: the 18 cancer spectra were taken as a reference library. The closeness or distance of each index spectrum from the cancer library was measured by calculating the Pearson correlation coefficient between the index spectrum and the spectrum of each member of the cancer library. For each index noncancer spectrum, the Pearson correlation of that spectrum with each of the 18 spectra in the cancer library was calculated, yielding 18 correlation values. The mean of these 18 correlations was then calculated as distance-C. For an index cancer spectrum, only 17 spectra in the library were used so that a spectrum was not compared to itself. For each index cancer spectrum, 17 correlation values were averaged to yield distance-C. Thus, every spectrum has a single value of distance-C.

Another distance measured, distance-A, is the usual straight line distance between two points in a multidimensional space. Across the 743 frequencies from 1503 to 761 cm^{-1} , distance-A was calculated as the square root of the sum of squared differences of normalized absorbances between two spectra. For a particular spectrum, distance-A was averaged across all spectra in the library. Differences between cancer and noncancer (spectral) groups on distance-C and distance-A were evaluated by t -tests. The difference between groups with respect to the distance-C measure was assessed by a simulation study or permutation test. In the latter test, 18 of the 47 patients randomly were designated as having cancer and placed in a library. This random

designation was performed 10,000 times and a t -statistic was calculated for each of the 10,000 trials. The proportion of simulated t -statistics that exceeded the obtained t -statistic yielded an empirically derived P value that served to check on the obtained P value.

The location and normalized absorbances of various peaks were examined. Four bands are especially obvious in most spectra. These occur at about 1650-, 1410-, 1230-, and 1060- cm^{-1} . Even though the 1650- cm^{-1} band was not within the frequency range of 1503 to 761 cm^{-1} , it was included in this comparison because it is an easily identifiable peak.

We wanted to determine if the distance from the cancer library was systematically related to the shape of the spectrum. Under the null hypothesis, cancer and noncancer spectra differ only randomly. That is, at a given frequency, a noncancer DNA spectrum is just as likely to occur above as below a cancer spectrum. Also, under this hypothesis, noncancer spectra that are more distant from the cancer library would be distant in a random fashion, sometimes lying above or below the mean cancer spectrum and lacking a systematic difference. Correlation was used to test this null hypothesis for the noncancer group; each spectrum had only one value of distance- C , but only one per integer wavenumber. If the normalized absorbances shifted in a systematic way with distance from the cancer library, then this relationship would be expected to be reflected in a correlation between distance and normalized absorbance. If there was a systematic shift, either a large positive or large negative correlation of the normalized absorbance across the data for the 29 noncancer patients would be expected when their normalized absorbances at a specific wavenumber were plotted versus distance- C . Wavenumbers that had substantial

correlation values between normalized absorbance and distance-C per point regions of the spectrum that shift in a systematic way. Thus, we calculated the Pearson correlations between distance-C and the normalized absorbance for each integer wavenumber. The same value of distance-C was used for a particular spectrum whenever it entered into a correlation calculation, but the normalized absorbance varied according to the specific wavenumber for which correlation was being calculated. In the range 1503 to 761 cm^{-1} , 743 Pearson correlation coefficients were calculated. The statistical significance of the correlation at each wavenumber was also assessed, and the Schweder and Spjøtvoll (24) method was used to estimate the number of frequencies for which there was a true association between distance from the cancer library and normalized absorbance.

Factor analysis was employed to study spectral variations and the relation of these variations to cancer versus noncancer DNA status. The form approach used was equivalent to principal component analysis. Analysis was performed on the normalized absorbance data without the mean removed, so that the first factor mainly reflects a mean spectrum across all cancer and noncancer DNA. The mean of the spectra for multiple sections from noncancer patients was taken to obtain a single spectrum DNA sample.

Analyses were conducted to relate spectral characteristics among cancer and noncancer DNA to $\bullet\text{OH}$ -induced base structural changes occurring in breast carcinogenesis as described previously (10). In the earlier study, striking associations were found between cancer/noncancer status and the ratio of ring-opening base structures to 8-OH-adducts. Thus, in the present study,

statistical model based on \log_{10} (Fapy-A/8-OH-Ade), having a high sensitivity (91%) and specificity (96%), was used (sensitivity is defined as the percentage of cancer patients correctly classified and specificity is defined as the percentage of noncancer patients correctly classified). This logarithm of concentration ratios was used for 29 patients, 10 with and 19 without cancer. The Pearson correlation was calculated between the Fapy A/8-OH-Ade model and the spectral descriptive characteristics. For patients with multiple sections, the mean of the \log_{10} ratios was used.

A model for classifying the status of patients with and without cancer was developed using logistic regression (25). Logistic regression yields a model for the probability that a patient is in the cancer group as a function of the descriptive characteristics included in the model.

B. RESULTS

1. Studies on Medaka Using Reduced Tissue Sample Weights

Preliminary findings indicate that sufficient DNA is obtained from a single Medaka liver to allow analysis by GC-MS/SIM in duplicate (Table 1). The modifications made in the DNA extraction procedure, the increased sensitivity of the Hewlett Packard instrument and other modifications in technique have made the "microtization" a reality.

Table 1. *Improved DNA Extraction Methodology*

Sample ID	Liver Weight (mg)	AMT. DNA (ug)*	A260/A280
EE3-93-027-17-14	3.9	45.7	1.65
EE3-93-027-17-5	6.4	92.5	1.82
EE3-93-027-17-11	3.0	40.5	1.90
EE3-93-027-18-24	3.0	63.8	1.91
EE3-93-027-18-27	3.4	68.0	1.84
EE3-93-027-18-33	4.1	93.8	1.86
EE3-93-027-17-26	3.6	71.0	1.91
EE3-93-027-17-6	5.7	30.0	1.63
EE3-93-027-17-19	5.2	136.0	1.68
EE3-93-027-17-20	2.2	40.5	1.80
EE3-93-027-17-28	3.6	65.1	1.78
EE3-93-027-17-29	5.8	44.6	1.75
EE3-93-027-17-47	3.3	30.0	1.73
EE3-93-027-17-23	6.3	60.8	1.66
EE3-93-027-17-33	5.4	67.3	1.83

* The significant differences obtained in yields are likely attributable to imprecision in "cut-points" in the solvent extraction.

2. Application of the GC-MS/SIM Technique to the Analysis of Human Normal and Cancerous

Female Breast Tissues

As a consequence of the work done on this project, substantial •OH-induced base lesions were found in the DNA of IDC of the female breast. However, virtually no information was available regarding relationships between the different base lesions in the normal and cancerous

breast. Such information is essential in understanding initial stages in the development of breast cancer and the potential of the base lesions as early predictors of cancer risk.

The •OH-induced DNA base lesions in normal RMT were compared to those from IDC and nearby MNT. Comparisons were then undertaken on relationships between the base lesion profiles in the normal and cancerous breast using twenty-two statistical models.

DNA from the RMT was characterized by a high ratio of ring-opening products (e.g., 4,6-diamino-5-formamidopyrimidine) to OH-adducts of adenine and guanine. A dramatic shift in this relationship in favor of carcinogenic hydroxy-adducts (e.g., 8-OH-Gua) was found in the cancerous breast. Statistical models with a high sensitivity (91%) and specificity (97%) provided a consistent means of classifying tissues (e.g., 96% correct).

The dramatic shift in the DNA base lesion relationships in oncogenesis is attributed to alterations in the redox potential of the breast favoring oxidative conditions and cancer formation. These findings suggest that base lesion profiles are potential sentinels for cancer risk assessment. Further, intervention in controlling the tissue redox potential may provide benefit in delaying or preventing early oncogenic changes and the ultimate manifestation of cancer.

Graphical analysis showed the logarithm of values to be more closely related to cancer vs. noncancer origin of tissue sections and more normally distributed than values on the natural scale. Thus, we used \log_{10} concentrations and \log_{10} ratios of concentrations in all analyses. The concentration of HMUra was below the detection limit of 0.0002 nmol/mg DNA for 14 sections and these sections were assigned a value of 0.0001 nmol/mg DNA. The mean values for cancer and noncancer tissue of the \log_{10} concentrations and ratios and the statistical significance of

differences were calculated using methods developed by Laird and Ware that, in our case, take account of the dependence of multiple sections from individual patients (21). The method is similar to ordinary multiple linear regression in other regards.

In order to build a model for predicting the origin of the tissue sections (cancer vs. noncancer), we used an extension of these methods for binary variables. In our context, this is a model for the probability that a specific tissue derives from a cancer or a noncancer patient. The probability is expressed as a function of \log_{10} concentrations or ratios of concentrations. To use it as a predictive model, a cut-off probability, P_C , is selected (e.g., $P_C = 0.5$) and tissue samples with an estimated probability above this value are labeled as cancer-derived. We calculated the sensitivity and the specificity of the classification, based on trial cut-off values from $P_C = 0.1$ to $P_C = 0.9$ in 0.1 increments, and chose the value of P_C that gave the highest combined values (expressed as a sum) of sensitivity and specificity.

We determined if mean concentrations or ratios of concentrations differed between MNT and IDC tissue from cancer patients, using the Laird-Ware model with the \log_{10} values as dependent variables, MNT vs. IDC as a dichotomous independent variable, and patient as a random effect.

The GC-MS/SIM analyses revealed dramatic differences in the concentrations of the DNA base lesions between the cancerous breast and the RMT. Both the IDC and the MNT were characterized by relatively high proportions of OH-adducts produced via the oxidation of the nucleotide bases. The base lesions were 8-OH-Ade, 8-OH-Gua and HMUra. Fapy derivatives (Fapy-A and Fapy-G), which are produced through reductive pathways from the initially formed

8-oxyl derivatives, were present in relatively small concentrations in the cancerous breast. However, the relationship between the concentrations of the OH-adduct and Fapy derivatives was dramatically different in the RMT. Overall, a clear distinction was evident between the ratios of Fapy:8-OH base lesion concentrations in the cancerous tissue and those of the normal tissue.

None of the tissues examined showed evidence of inflammatory responses during histologic examination. Thus, there is no evidence for any contribution from infiltrating cells in the proportions of reported DNA lesions. Each of the IDC and MNT specimens had a mirror-image "control" histologic section prepared and examined in the absence of any knowledge of the DNA base lesion data.

Fapy-A concentrations predominated in the RMT sections compared to 8-OH-Gua by a factor of ~ 4- to 10-fold; as depicted in Figure 1 below:

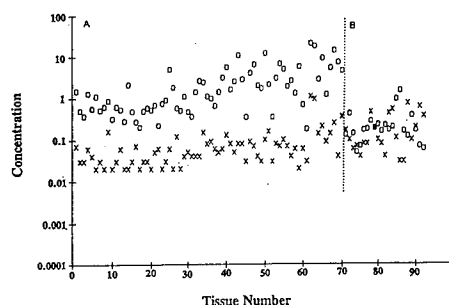


Figure 1. A scatterplot depicting the relationship between the \log_{10} concentration of the base lesions (nmol/mg DNA) versus tissues analyzed. Panel A: RMT; panel B: IDC and MNT. Circles represent Fapy-A and X represents 8-OH-Gua.

Remarkably high concentrations of Fapy-A were found in the RMT (mean \pm S.E. = 2.9 ± 0.49 nmol/mg DNA; one base lesion in 320 normal bases). Surprisingly, for example, one

patient had a RMT section that contained 21.0 nmol Fapy-A/mg DNA, or one base lesion in 46 normal bases. Overall, high concentrations of Fapy derivatives in the RMT did not prevent the formation of significant concentrations of OH-adducts in some tissues. The tissue section from the patient mentioned above had a relatively high 8-OH-Gua concentration of 1.1 nmol/mg DNA (one base lesion in 540 normal bases). Thus, two aspects relevant to \bullet OH-induced carcinogenesis are the redox status of the tissue and the absolute concentrations of mutagenic OH-adducts (e.g., 8-OH-Gua). Both of these parameters would be pivotal in the assessment of carcinogenic risk factors.

In contrast, the IDC and MNT sections were characterized overall by elevations in 8-OH-Gua compared to RMT, coupled with a marked depletion of Fapy-A residues. Thus, the results further indicate that fundamental differences exist in the nature of the \bullet OH-induced base damage in relation to cancerous and noncancerous tissues. This is evident from the histogram shown below in Figure 2, which depicts the concentrations (mean \pm S.E.) of the base lesions in the RMT, MNT and IDC tissues.

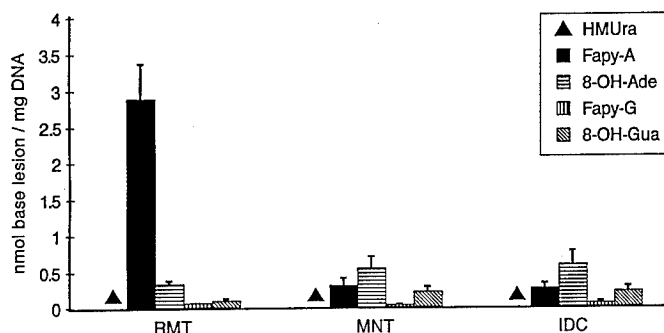


Figure 2. A histogram of DNA base lesion values (mean \pm standard error) for the cancerous (MNT and IDC) and normal female breast (RMT). The relatively low base lesion values for HMUra are designated by a triangle (\blacktriangle): RMT = 0.0007 \pm 0.0001; MNT = 0.0021 \pm 0.0004; and IDC = 0.0021 \pm 0.0005.

A statistical analysis of the data was conducted yielding the mean values of the various indicators (\log_{10} concentrations or \log_{10} ratios of concentrations) for cancer and noncancer tissue and the statistical significance of differences using the Laird-Ware regression model. The sensitivity and specificity were calculated using the predictive logistic regression model. Most significance levels are strikingly small, indicating prominent differences between cancer and noncancer tissue with respect to a wide array of predictors. No correlation between patient age and predictors was observed. Consequently, age was not included in the analyses, neither in the cancer dataset nor in the noncancer dataset, and can be ruled out as a cause of the strong association between the base lesions and the origins of tissue sections. Examples of these data are presented in Table 2.

Due to the number of comparisons made (22 predictors were assessed; 9 of high sensitivity and specificity are given as examples in the tabular data above), it is likely that one or two would be statistically significant by chance alone. However, if all the p-values determined are multiplied by 22, which is the conservative Bonferroni adjustment, almost all of the p-values would still be statistically significant, including those for the \log_{10} ratio that is considered further below.

Table 2. Mean \log_{10} concentrations and \log_{10} ratios of concentrations of DNA-base lesions from reductive mamoplasty tissue (RMT) and cancerous breast tissue (IDC and MNT) and sensitivity and specificity based on statistical models.*

Indicator	Non-cancer patients			Cancer patients		Predictive Model (Logistic regression)		
	P-value†	Mean§	S.E.	Mean§	S.E.	Sensitivity y (%)	Specificity (%)	P-value‡
<i>log₁₀</i> (concentrations)								
HMUra	.0000	-3.3	.1	-2.8	.1	91	69	.0001
Fapy-A	.0000	.2	.1	-.7	.1	82	93	.0000
<i>log₁₀</i> (ratio of conc.)								
Fapy-A/HMUra	.0000	3.6	.1	2.0	.1	91	97	.0001
Fapy-A-8/OH-Ade	.0000	.9	.1	-.3	.1	91	96	.0000
Fapy-A/8-OH-Gua	.0000	1.4	.1	.2	.1	91	94	.0001
Fapy-A/(8-OH-Ade + 8-OH-Gua)	.0000	.8	.1	-.4	.1	91	97	.0001
Fapy-A/(8-OH-Ade + 8-OH-Gua + HMUra)	.0000	.8	.1	-.4	.1	91	97	.0001
(Fapy-A + Fapy-G)/ (8-OH-Ade + 8-OH- Gua)	.0000	.8	.1	-.4	.1	91	97	.0000
(Fapy-A + Fapy-G)/ (8-OH-Ade + 8-OH- Gua + HMUra)	.0000	.8	.1	-.4	.1	91	97	.0000

*All analyses are based on n = 15 cancer patients with a total of 22 tissue samples and 15 non-cancer patients with a total of 70 tissue samples.

†Based on linear regression random-effects model, testing the null hypothesis of equality of means for cancer and non-cancer (RMT) patients.

§Mean of the \log_{10} values for the tissue specimens.

‡Based on logistic regression random-effects model, testing the null hypothesis that the \log_{10} value is not associated with cancer vs. non-cancer (RMT) classification of tissue section.

We used the model for the \log_{10} ratio of summed Fapy derivatives to summed OH-adducts plus HMUra because it is based on reductive vs. oxidative conversion pathways of the initial 8-oxyl derivative and is one of the best models for predicting the cancer vs. noncancer origin of tissue. However, as the above table clearly shows, there are other models with high sensitivity and specificity and very small significance levels. The size of this dataset does not allow definitive selection among the several good models. The predictive equation is :

$$\log_e [P/(1-P)] = 0.76 - 6.34 \times \log_{10} (\text{ratio})$$

where P is the probability that a tissue sample derives from a cancer patient and "ratio" refers to the ratio of the sum of the two Fapy derivatives to the sum of the two OH-adducts plus HMUra. The standard errors of the constant term and for the multiplier of the \log_{10} ratio in the model above are 0.58 and 1.53, respectively. Using the model and the cut-off $P_c = 0.5$, tissue samples with an estimated probability $P > 0.5$ were classified as cancer-derived, and those with $P \leq 0.5$ were classified as noncancer derived. The corresponding ratio of concentrations that best divides cancer from noncancer samples is 1.32. As can be seen in Table 2, the sensitivity (91%) and specificity (97%) are both very high.

In addition to the classification based on (Fapy-A + Fapy-G)/(8-OH-Ade + 8-OH-Gua + HMUra), we also show the classification based on a model of high sensitivity and specificity using the ratio (Fapy-A/8-OH-Gua). In the latter model, the predictive equation is:

$$\log_e [P/(1-P)] = 3.71 - 5.51 \log_{10} (\text{ratio})$$

The standard errors of the intercept and multiplier of \log_{10} are 1.20 and 1.38, respectively. The cut-point for the predictive probability used in classifying a cancer-derived tissue is $P_c > 0.4$, which corresponds to a ratio of concentrations of 5.6 or less.

The comparison of \log_{10} concentrations and ratios between IDC and MNT showed no statistical differences, thus indicating that the observed DNA base modifications were pervasive in both the IDC and MNT. However, due to the small sample size ($n = 22$ sections), large differences in concentrations or ratios between IDC and MNT cannot be ruled out.

Based upon the pronounced differences in base lesion profiles and concentrations between the cancer and noncancer tissue, a graph of predicted probability of the cancerous origin of a tissue vs. \log_{10} of the concentration ratios was constructed. This demonstrates the strong ability of this model to discriminate the nature of each tissue. The data are given below in Figure 3:

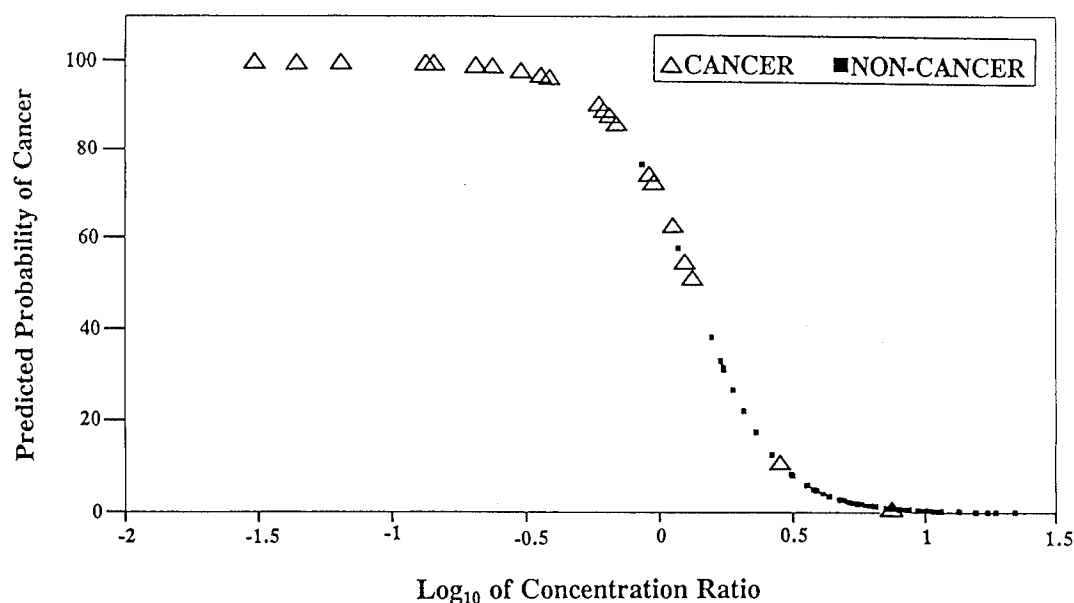


Figure 3. The predicted probability of the cancerous origin of a tissues is plotted with \log_{10} of the concentration ratio (Fapy-A + Fapy-G/(8-OH-Ade + 8-OH-Gua + HMUra) for all samples analyzed.

It is known that oxidative stress is linked to cancer formation and that increases in OH-adducts (e.g., 8-OH-Gua) are a likely consequence of oxidative conditions in the cell. Consistent with this is the concept that the oxidative modifications of DNA structure reported in breast cancer are the probable basis for the carcinogenic action of H_2O_2 generation. Although multiple biochemical processes may be involved, it is suggested that the $\bullet OH$ may arise as a consequence of the formation of H_2O_2 from redox cycling of endogenous (e.g., hormones) or exogenous effectors (e.g., polychlorinated biphenyls [PCBs] and chlorinated hydrocarbons), mediated by cytochrome P-450 and cytochrome P-450 reductase (27).

It is noteworthy that breast tissues of women with breast cancer have elevated concentrations of PCBs compared to those with benign breast disease (28). In this regard, the previously reported (29) relationship between fat intake and HMUra in DNA of peripheral nucleated blood cells of women with breast cancer may reflect, at least in part, the influence of organic xenobiotics enriched in the dietary fat.

The H_2O_2 , which is readily transported across the nuclear membrane, is likely converted to the $\bullet OH$ via the Fe^{++} -catalyzed Fenton reaction. The subsequent attack of the $\bullet OH$ on the nucleotide bases results in the formation of the 8-oxyl derivatives of the purines and the hydroxylation of thymine to form HMUra. At this point, the conversions of the purines can either lead to oxidatively-formed OH-adducts that potentially increase cancer risk or to reductively-formed Fapy derivatives that are putatively non-genotoxic. The synthesis of the ring-opening structures appears to protect the DNA from potentially mutagenic OH-adduct formation and, as such, reflects a unique antioxidant role for the DNA base structure. Strikingly, the nature of these transformations occurring in the cell leading to differing classes of base

lesions is entirely consistent with the redox-coupled pathways of $\bullet\text{OH}$ -induced purine modifications occurring in aqueous solution as described by Steenken (30). It is particularly noteworthy that oxidants (e.g., O_2) in aqueous solution quantitatively suppress Fapy derivatives and increase the yield of 8-OH-adducts. In view of this, we were not surprised that the most effective predictive models shown in Table 2 [e.g., (Fapy-A + Fapy-G/8-OH-Ade + 8-OH-Gua + HMUra)] were completely consistent with the above-mentioned pathways. The proposed pathway for the synthesis of the OH-adducts and Fapy derivatives is given below in Figure 4:

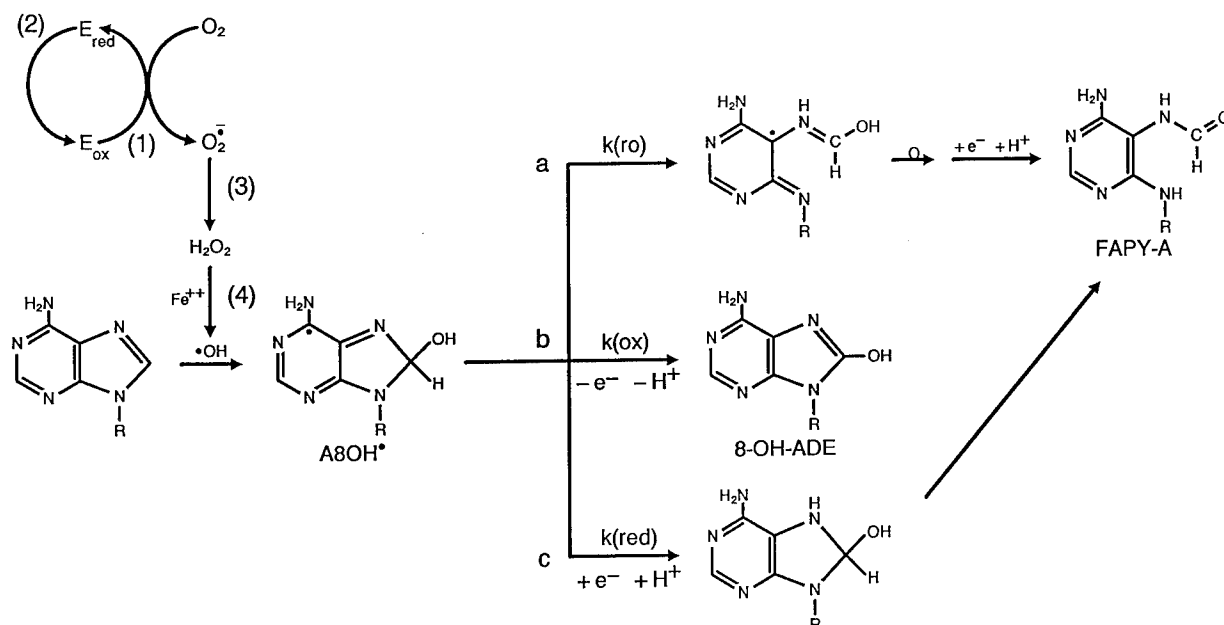


Figure 4. A proposed scheme for the formation of the ring-opening (Fapy) derivatives and 8-OH-adducts in the female breast. As an example, adenine is converted to the 8-oxyl derivative ($\text{A8OH}\bullet$) via the attack of the $\bullet\text{OH}$. The $\bullet\text{OH}$ is derived from the Fe^{++} -catalyzed conversion of H_2O_2 (pathway 4). The H_2O_2 may arise from multiple metabolic processes occurring in the breast epithelial cells, one of which may include the redox cycling of an endogenous or exogenous effector molecule (E) via cytochrome P-450 oxidase (pathway 2) and cytochrome P-450 reductase (pathway 1). The $\text{A8OH}\bullet$ can be converted oxidatively to 8-OH-Ade (pathway b) or reductively to Fapy-A (pathway a or c). The redox balance in the breast cells would dictate the ratio, for example, of 8-OH-Gua:Fapy-A formed with increases in cellular oxidants favoring pathway b and potential cancer formation. The cytochrome P-450 pathways are essentially as described by Deodatta et al.¹⁹ and the aqueous solution redox chemistry and transformation reactions are based on those described by Steenken.

We conclude that the •OH-induced oxidative base damage likely represents an event of considerable importance in the early development of breast cancer. For example, the DNA from several sections of the normal breast contained greater than one 8-OH-Gua base lesion in 1,000 normal bases. The presence of elevated levels of 8-OH-Gua in the DNA of a relatively small number of normal breast sections is perhaps to be anticipated considering the fact that one out of eight women develop breast cancer on a lifetime basis. In this context, the attack of the •OH on the base structure of the breast DNA would be expected to result in the activation or augmentation of nuclear oncogenes and the deregulation of tumor suppresser genes, such as p53 (31). Other genotoxic changes are likely and the greater the intensity of the radical attack, the greater the expectation of mutagenic events occurring.

In considering the proposed role played by cellular redox conditions and base lesion formation in the etiology of breast cancer, it was recognized that DNA repair may potentially play a significant part in processes that govern these circumstances. Enzymes capable of repairing Fapy and 8-hydroxypurine derivatives are known to be constitutively expressed in *E. coli* and mammals (32). Moreover, growing evidence indicates that one of these enzymes, the FGP protein, is involved in the repair of both Fapy and 8-hydroxy base lesions (33). Although the 8-hydroxy-dG derivative may result in some inhibition of DNA replication, more specifically it is known to be mutagenic, resulting in miscoding lesions due to a 1-to-2% level of misrepair (34). However, there is no current evidence supporting a mutagenic property for the ring-opening lesions. Instead, the Fapy residues have been shown to block DNA synthesis (17). Thus, unrepaired Fapy residues, which are abundant in the DNA from the normal breast, would not be expected to be genotoxic, although they may be cytotoxic. For differential DNA repair to explain the present findings, the

transition between high ratios of Fapy : hydroxy derivatives in the RMT to low ratios in the IDC and MNT would be expected, for example, to involve preferential repair of the Fapy-A residue while the 8-hydroxy derivatives increased. This circumstance does not conform to the known behavior of the DNA repair mechanisms involved. In fact, support for our hypothesis for oxidation-driven base lesion changes during oncogenesis in breast cancer includes evidence for decreased DNA repair in cancers of the breast, colon and lung (35), the presently demonstrated increased concentrations of OH-adducts in the cancerous breast, and the finding that trans-tamoxifen exerts an antioxidant effect (i.e., a decrease in tumor promoter-induced H_2O_2 formation in human neutrophils) that correlates with diminished concentrations of oxidatively-formed HMUra. Of additional significance is the fact that patients with a single breast cancer are at increased risk of having a second primary tumor in the breast (36). Our findings showing that \log_{10} base concentrations and ratios between IDC tissue and MNT were not statistically different is consistent with this finding. That is, significant oxidative base damage in the DNA would be expected to still be present in the MNT after the tumor is removed, thus potentially increasing the risk of a second tumor occurring.

Regarding the statistical models, the sensitivity and specificity calculated from our specific dataset (Table 2) can be expected to be somewhat high (non-conservative) compared to the specificity and sensitivity that would be calculated from a trial of the predictive equation with a new population of tissue samples. This bias occurs because the sensitivity and specificity have been optimized within this specific dataset. The statistical significance calculations, however, are unbiased for inference about a similar mix of RMT and cancer patients as observed here.

Thus, the promise of this area is very strong (based on significance levels), but specific screening models should be based on a larger dataset with more cancer patients and normal individuals.

The sensitivity and specificity in this study have been calculated for classification of tissue samples and not for classification of individual patients. However, multiple tissue samples from patients are very likely to be classified consistently. For example, classification of tissue sections based on $\log_{10}(\text{Fapy-A} + \text{Fapy-G})/(\text{8-OH-Ade} + \text{8-OH-Gua} + \text{HMUra})$ and $\log_{10}(\text{Fapy-A}/\text{8-OH-Gua})$ showed only 4/92 and 6/92 incorrect classifications of cancer vs. noncancer tissue, respectively (Table 3). Thus, the method is most promising for use in classification of patients based on individual samples of tissue.

Table 3. *Classification of tissue sections using predictive model based on ratio of concentrations: (Fapy-A + Fapy-G)/(8-OH-Ade + 8-OH-Gua) or (Fapy-A)/(8-OH-Gua)*.*

Patient #, <u>type of patient</u>	Classification of sections based on (Fapy-A + Fapy-G)/(8-OH-Ade + 8-OH-Gua + HMUra)	Classification of sections based on (Fapy-A)/(8-OH-Gua)
	<u>N incorrect/N total</u>	<u>N incorrect/N total</u>
1, RMT	0/7	0/7
2, RMT	0/13	2/13
3, RMT	0/4	1/4
4, RMT	0/5	1/5
5, RMT	1/5	0/5
6, RMT	0/5	0/5
7, RMT	0/5	0/5
8, RMT	0/6	0/6
9, RMT	0/5	0/5
10, RMT	0/2	0/2
11, RMT	0/2	0/2
12, RMT	0/2	0/2
13, RMT	0/2	0/2
14, RMT	1/5	0/5
15, RMT	0/2	0/2
16, cancer	0/1	0/1
17, cancer	0/1	0/1
18, cancer	0/1	0/1
19, cancer	0/1	0/1
20, cancer	0/1	0/1
21, cancer	0/1	0/1

Table 3 cont'd...

Patient #, <u>type of patient</u>	Classification of sections based on (Fapy-A + Fapy-G)/(8-OH-Ade + 8-OH-Gua + HMUra) <u>N incorrect/N total</u>	Classification of sections based on (Fapy-A)/(8-OH-Gua) <u>N incorrect/N total</u>
22, cancer	0/1	0/1
23, cancer	0/1	0/1
24, cancer	0/2**	0/2**
25, cancer	0/2**	0/2**
26, cancer	0/2**	0/2**
27, cancer	2/2**	2/2**
28, cancer	0/2**	0/2**
29, cancer	0/2**	0/2**
30, cancer	0/2**	0/2**
Total	4/92	6/92

*Tissue was classified as derived from a cancer patient if ratio of concentrations (Fapy-A + Fapy-G)/(8-OH-Ade + 8-OH-Gua + HMUra) < 1.32 (second column) or (Fapy-A)/(8-OH-Gua) < 5.6 (third column).

**Represents paired IDC and MNT from the same patient.

The models considered were based on retrospective analysis of the origin of the tissue. The curve displaying the probability of cancer vs. the \log_{10} of base lesion concentration ratios clearly affirms the difference in the results between the cancer and noncancer patients (See Figure 3). Given the biological implications of the differing classes of base lesions, which are formed as a function of the cellular redox potential as discussed above, it is reasonable to conclude that this may represent a basis for prospective cancer risk to be estimated through log transformations of base lesion concentrations in the DNA of breast tissues. In this regard, it is noteworthy that the probability model classifies certain of the tissues examined as having base lesion concentrations that may reflect transitional states between those of normal and cancerous tissue. Evaluation of the potential risk of an individual developing breast cancer at an early stage represents an important potential of this analysis.

The model that predicts cancer vs. noncancer status may thus also predict future risk as well. Evaluation of the present methodology for clinical application would require a prospective study of women at variable predicted risks for developing breast cancer, based on models such as we have developed here. Such a study would naturally include an evaluation of relationships between diet, ethnic differences, reproductive history, familial history, and other relevant factors. If prospective studies confirm our results, individuals identified to have a heightened predicted cancer risk would be expected to benefit from close monitoring and possible intervention with antioxidants or other agents. The close association between cancer chemoprotection and compounds with antioxidant activity (37-40) is consistent with this potential and the results presented in this paper.

In conclusion, it is clear that the DNA base lesion profiles reflect intrinsic differences that exist between normal and cancer-derived tissues in a manner regulated by the redox condition of the breast cells. The nature of these base lesions present in the tissues represents a useful sentinel for evaluating the prevailing redox conditions. Further, potential mutagenic damage to the DNA base structure can be assessed. On this basis, it is a logical assumption that a shift in the base profiles characteristic of normal breast tissue to profiles characteristic of cancer tissue is early evidence for a heightened risk of cancer formation. In this context, the results presented describe a potentially powerful method for defining characteristic changes in the DNA of female breast tissue during oncogenesis. Given the fact that analyses can be performed readily on small amounts of biopsied tissue, this method could ultimately have wide application for determining individuals at risk in the population.

3. Application of FT-IR Techniques to the Analysis of Human Normal and Cancerous Female Breast Tissues

Alterations in the FT-IR spectra between normal and cancerous female breast tissues are significant in a number of areas of the spectrum investigated. To illustrate, the mean DNA spectrum ($2000\text{-}700\text{ cm}^{-1}$) for patients with and without cancer is shown in Fig. 5. The main features of the spectral profiles resemble those of DNA from normal tissues obtained previously in studies using IR spectroscopy (19, 41). For example, the area $1700\text{ to }1500\text{-cm}^{-1}$ was assigned to strong CO stretching and NH_2 bending vibrations, and $1550\text{-}1300\text{ cm}^{-1}$ was assigned to weak NH vibrations and CH in-plane base deformations; 1240 cm^{-1} represents medium PO_2 antisymmetric-stretching vibrations of the phosphodiester backbone; $1100\text{ to }900\text{ cm}^{-1}$ represents strong PO_2 symmetric stretching vibrations ($\sim 1080\text{ cm}^{-1}$), the CO-stretching vibrations of the deoxyribose moiety and the PO-stretching vibrations of the PO group of the phosphodiester backbone. The lower portion of Fig. 5 illustrates the P values associated with tests for differences between cancer and noncancer spectra at each wavenumber. A large P value suggests that there was no difference between cancer and noncancer spectra, whereas a small P value indicates that the P value was significant. The unequal variance version of the t -test was used because patients without cancer had more diverse spectra than did those with cancer. As shown in Fig. 5, a number of frequency areas have P values less than 0.05. Among the 743 P values from $1503\text{ to }761\text{ cm}^{-1}$, the Schweder and Spøtjvoll (28) method suggested that the null hypothesis was likely to be false at approximately 300 frequencies. That is, the difference between cancer and noncancer spectra was real at approximately 300 frequencies. The spectra were not uniform in the occurrence of cancer and noncancer differences. The P values on the

right side of Fig. 5 are generally smaller than those on the left side and most small P values occurred at 1200 cm^{-1} or lower. The small P values identified areas in which cancer and noncancer spectra were notably different. These include approximately the 1720 -, 1170 -, 1070 -, 1030 -, 940 -, 860 - and 790-cm^{-1} areas.

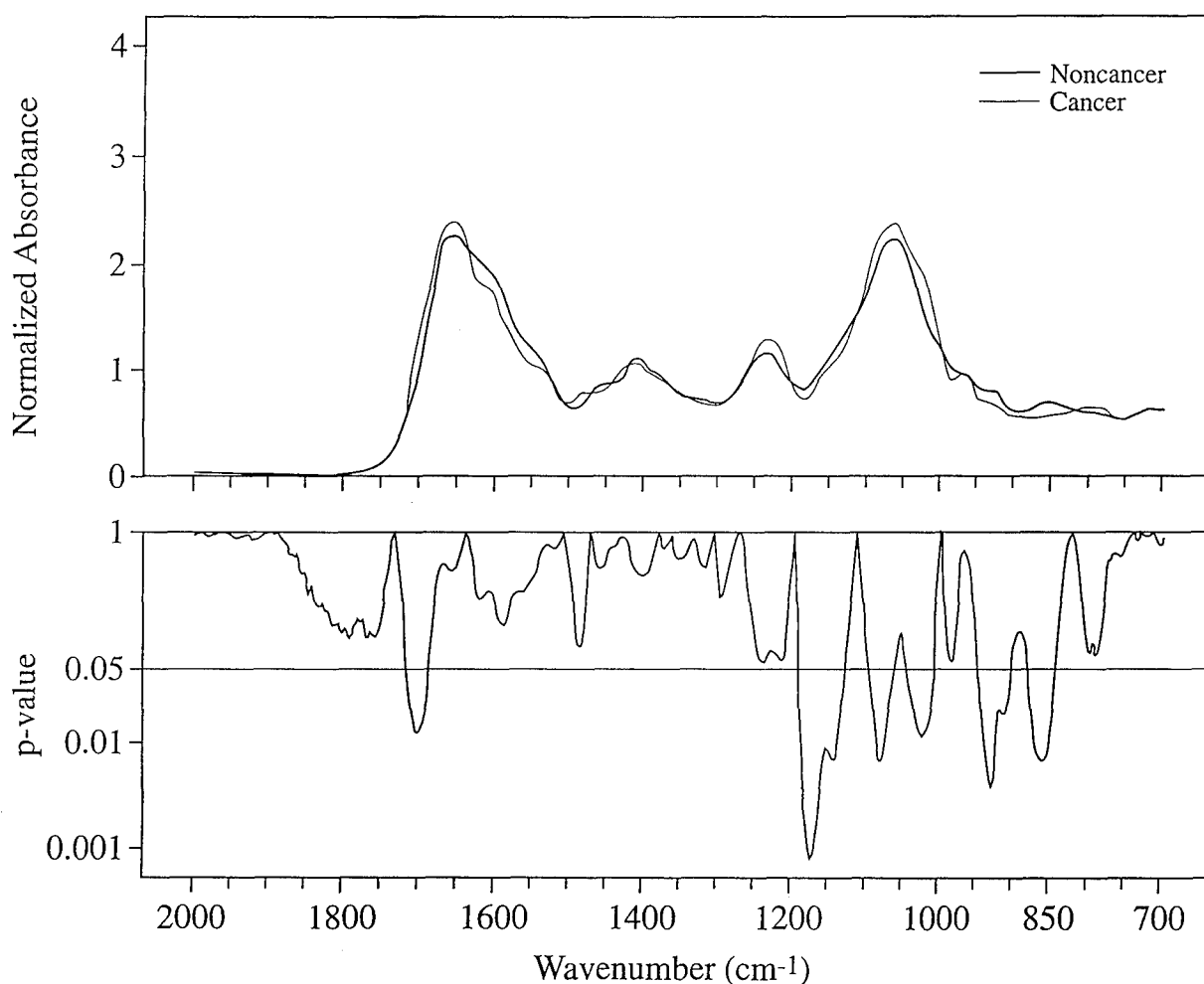


Figure 5. Mean normalized absorbance spectra of patients with cancer ($n=18$) and without cancer ($n=29$) and the statistical significance of cancer versus noncancer absorbances, based on the t-test with unequal variances. P values were not adjusted for multiple testing (mean spectra, top; P values, bottom).

Using distance-C, based on the mean Pearson correlation, a global test that is not influenced by the problem of multiple P values also was used to assess the difference between cancer and noncancer spectra. For the primary range of interest, $1503\text{--}761\text{ cm}^{-1}$, distance-C for patients with and without cancer was 0.93 and 0.87, respectively. Using the unpaired t -test with unequal variances, $P = 0.003$ for the cancer/noncancer difference. In addition, a permutation test yielded $P = 0.004$ for this difference. These results are shown in Table 4, together with tests for other spectral features including distance-A and spectral intensities and locations for the four distinctive peaks between 2000 and 700 cm^{-1} (peaks A, B, C, and D, reading from highest to lowest frequency).

It is notable that the noncancer spectra were more diverse than the cancer spectra and generally more dissimilar. Among the patients without cancer, 12 of 29 (41%) had distance-C lower than 0.9 (lower values indicate greater distance) whereas among the patients with cancer, only 2 of 18 (11%) had distance-C lower than 0.9.

Table 4. *Comparison of spectral descriptive characteristics for cancer (n=18) and noncancer (n=29) patients.*

Item	Cancer mean (\pm SD)	Noncancer mean (\pm SD)	P-value
Distance C: mean Pearson correlation with cancer library for $1503\text{--}761\text{ cm}^{-1}$	0.93 (± 0.03)	0.87(± 0.10)	0.003* [†]
Distance A:airline distance from cancer library for $1503\text{--}761\text{ cm}^{-1}$	5.4 (± 0.8)	7.7 (± 3.6)	0.01 [†]
Peak intensities and wavenumber location			
Peak A in 1652 cm^{-1} area			
Normalized absorbance	2.4 (± 0.5)	2.3 (± 0.5)	0.6
Location (cm^{-1})	1652 (± 6)	1652 (± 7)	0.9
Peak B in 1410 cm^{-1} area			
Normalized absorbance	1.1 (± 0.2)	1.2 (± 0.3)	0.4
Location (cm^{-1})	1412 (± 6)	1407 (± 10)	0.02 [†]

Table 4 cont'd...

Item	Cancer mean (\pm SD)	Noncancer mean (\pm SD)	P-value
Peak C in 1233 cm^{-1} area			
Normalized absorbance	1.3 (\pm 0.2)	1.2 (\pm 0.3)	0.09
Location (cm^{-1})	1232 (\pm 7)	1235 (\pm 7)	0.1
Peak D in 1061 cm^{-1} area			
Normalized absorbance	2.4 (\pm 0.2)	2.3	
Location (cm^{-1})	1060 (\pm 6)	1062 (\pm 0.3) (\pm 13)	0.1 [†] 0.4 [†]
Log_{10} (Fapy-A/8-OH-Ade)	-0.31 (\pm 0.35) (n=10)	0.39 (\pm 0.61) (n=19)	0.001 [†]

SD: standard deviation

* P-value based on permutation test is 0.004.

[†] t-test based on unequal variances.

a. Systematic shifts in the spectra toward a cancer-like phenotype

The possibility was investigated that those noncancer spectra that are distant from the cancer library differ in a systematic rather than in a random way from the cancer spectra. The larger distances (smaller distance-C values between a given spectrum and the cancer library) can occur in one of two ways. First, the difference may be random in that at a given frequency, those spectra that are distant sometimes lie above and sometimes lie below the mean cancer library spectrum in a random fashion. Alternatively, at a given frequency the spectra with greater distances generally may lie below (or above) the mean cancer DNA spectrum. To investigate the possibility of a systematic shift in the spectrum with distance, the Pearson correlation of normalized absorbance with distance from the cancer library (expressed as distance-C) was calculated for each frequency. The results in Fig. 6 show some strikingly large correlations. If

the distance from the library was not associated with increasingly positive or negative deviations around the mean cancer profile, then most of the correlations in Fig. 6 would lie close to zero and between the upper and lower ($P = 0.05$) horizontal lines. Instead there were wide fluctuations in the correlations, with a number lying even beyond $r = \pm 0.6$, the horizontal line for $P = 0.001$. Using the Schweder and Spøtjvoll (24) method approximately 500 of the 743 frequencies of $1503\text{-}761\text{ cm}^{-1}$ were estimated to violate the null hypothesis of no association between absorbance and distance from the cancer library. The most significant correlation occurred at 1172 cm^{-1} , in which $r = -0.85$, and the unadjusted $P = 5 \times 10^{-9}$. Even multiplying this P value by 743 for a highly conservative Bonferroni correction for multiple testing (42) still yields a P value of 4×10^{-6} . Thus, the null hypothesis of no association between distance from the cancer library and spectral absorbance can be rejected decisively. Again, it is important to remember that spectra must converge necessarily on the cancer profile as the distance becomes smaller (distance-C approaches 1.0). However, if the process is random (the null hypothesis), there is no reason why spectra at a given frequency should approach the mean cancer library from one side rather than from randomly above or below it. Fig. 7 shows the consistent change of the noncancer spectrum for two frequencies. These are both in the area $1250\text{-}1000\text{ cm}^{-1}$ in which the cancer/noncancer differences are most striking, based on the t -test.

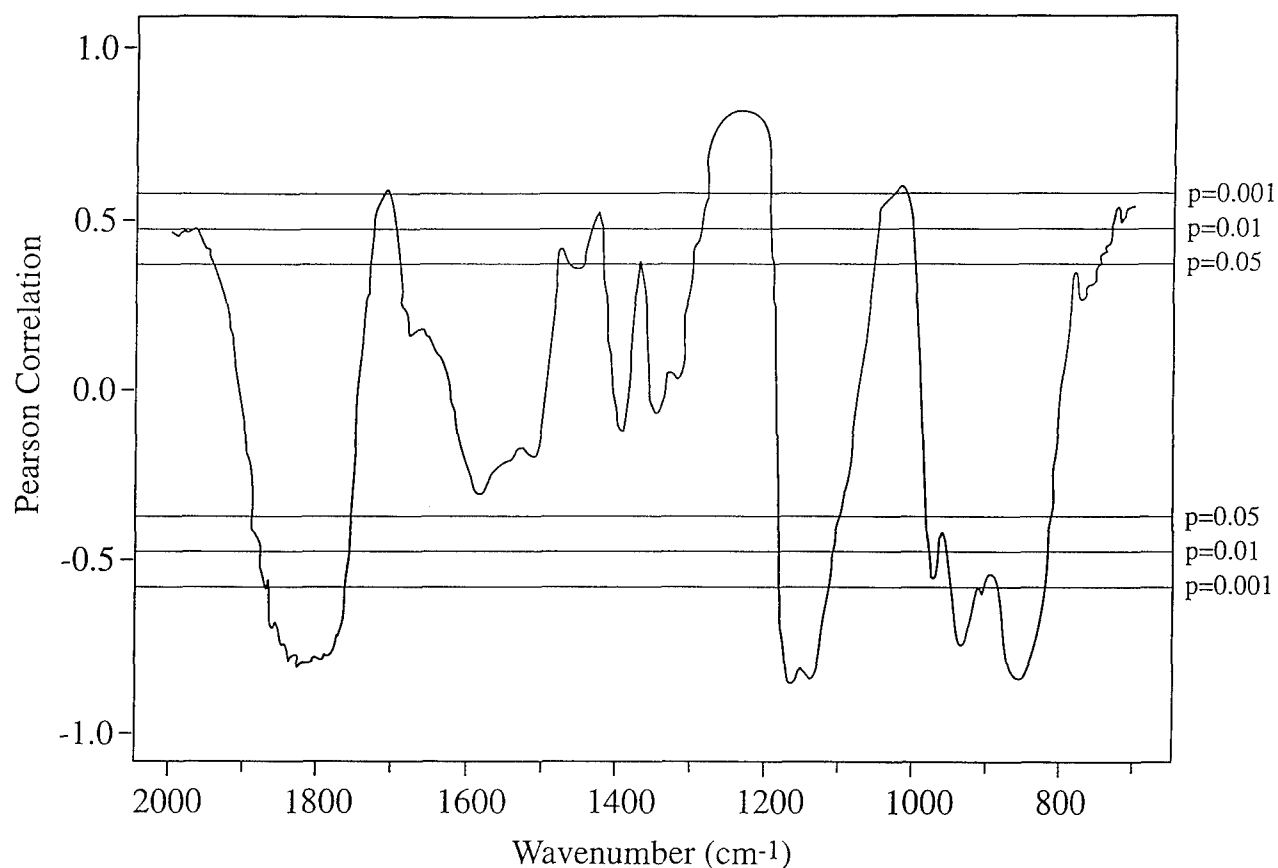


Figure 6. Pearson correlation coefficient between normalized absorbance and distance-C from cancer library. Horizontal lines indicate unadjusted P values. The smallest P value occurs at 1172 cm^{-1} in which the unadjusted $P = 5 \times 10^{-9}$.

To determine if consistent changes were also occurring within the cancer group, we used analyses similar to those presented earlier that showed consistent changes in the noncancer spectra in relative to distance from the cancer library. In this case, the noncancer spectra furthest from the cancer library served as a reference library. The analyses show that as distance between the cancers and this group of noncancers increases, some of the frequency areas show a trend of increasing or decreasing normalized absorbance, rather than random variation.

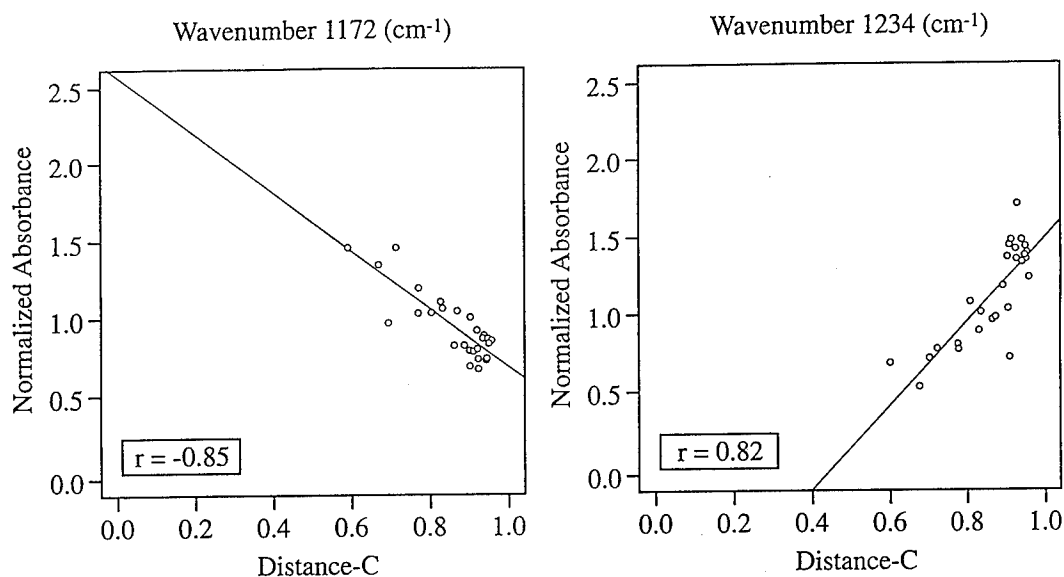


Figure 7. The relationship of the distance of patients without cancer from the cancer library (distance-C) and normalized absorbances at frequencies 1172 and 1234 cm^{-1} . Patients who were closer to the cancer library had an increasingly higher absorbance at 1234 cm^{-1} and an increasingly lower absorbance at 1172 cm^{-1} .

b. Relationship of spectral features to base modifications

Some of our spectral descriptive measures were related to the Fapy-A/8-OH-Ade base model (9). The logarithm of the ratio of base concentrations was correlated significantly with several of the spectral descriptive measures as shown in Table 5. This analysis has reduced significance compared to analyses based solely on spectral data because of the smaller number of patients with GC-MS/SIM data ($n = 10$ cancer; $n = 19$ noncancer). Still, several of the correlations were statistically significant or had correlations of ± 0.3 or larger. These correlations of independently derived spectral measures with a GC-MS-based measure of cancer risk mutually support the finding that both approaches are able to discriminate between patients with and without cancer and establish potential cancer risk levels.

Table 5. *Pearson correlation of spectral descriptive characteristics with \log_{10} ratio of concentrations of Fapy-A to 8-OH-Ade ($n=10$ cancer and $n=19$ noncancer patients).*

Item	Correlation	P-value
Distance C: Pearson correlation with cancer		
library for 1503-761 cm^{-1}	-0.53	0.003
Distance A: airline distance from cancer		
library for 1503-761 cm^{-1}	0.39	0.04
Peak intensities and wavenumber location		
Peak A in 1652 cm^{-1} area		
Normalized absorbance	0.22	0.3
Location (cm^{-1})	-0.22	0.3
Peak B in 1410 cm^{-1} area		
Normalized absorbance	0.33	0.08
Location (cm^{-1})	-0.59 (-0.48)*	0.0008 (0.008)*
Peak C in 1233 cm^{-1} area		
Normalized absorbance	-.45	0.01
Location (cm^{-1})	0.21	0.3
Peak D in 1061 cm^{-1} area		
Normalized absorbance	-.48	0.008
Location (cm^{-1})	0.57 (0.39)*	0.001 (0.03)*

* The Spearman correlation coefficient and its p-value are shown in parentheses when an outlier may have influenced the Pearson correlation coefficient.

c. Breast cancer risk model

The spectral descriptive measures were combined into a model for cancer risk assessment. Table 6 shows the model, based on logistic regression, along with the observed

sensitivity and specificity for this particular data set using the first six factors derived earlier. The association between the factors and cancer/noncancer status is significant ($P = 0.001$); the sensitivity and specificity are both 83%. These sensitivity and specificity values, however, are not unbiased because the cut-point value of predicted probability used to classify patients as having or not having cancer (Probability ≥ 0.4) was based on inspection of the predicted probabilities. Nevertheless, the statistical significance of the association of the six factors with cancer/noncancer status ($P = 0.001$) was unbiased. The predicted probability of cancer as a function of this model, which produces a risk score, is shown in Fig. 8. The plot shows that there was some overlap of patients with and without cancer in the middle range of predicted probabilities, whereas at the lowest and highest levels of predicted probability patients with and without cancer are separated clearly.

Table 6. *Predictive models for cancer vs. noncancer status based on a logistic regression analysis of first six factors from factor analysis.*

Principal Components	Model Coefficient	S.E.	P-value
Factor 1	19.0	8.8	0.03
Factor 2	0.01	0.24	1.0
Factor 3	6.8	3.0	0.02
Factor 4	3.6	1.7	0.04
Factor 5	-3.9	1.7	0.02
Factor 6	-1.5	0.6	0.02
Constant	568.3	263.5	0.03
Entire model			0.001*

* P-value for null hypothesis that the six principal components do not improve prediction of cancer/noncancer status.

† Sensitivity = 83% (15/18 cancer patients correctly classified); specificity = 83% (24/29 noncancer patients correctly classified). Based on logistic regression model using predicted probability $p \geq 0.4$ to classify patients into cancer group, used to define how the sensitivity was calculated.

The power of the predictive model was demonstrated by an independent test. The MNT nearby the tumor was analyzed to yield spectra in the same manner described earlier. The factor scores for these patients were calculated and used in the predictive model described above. Eight of 10 of the patients analyzed were classified as having cancer (predicted probability ≥ 0.4) based on the model. As indicated in Fig. 8, most of the MNTs were well above the 0.4 cut-point for predicted probability of cancer (one of the MNT specimens was from a patient who also provided a cancer specimen used to develop the predictive model; both specimens were classified as cancer by the model). In a practical setting, such models would have to be developed using a larger sample size and then tested on a new panel of patients whose specimens were not used for development of the model.

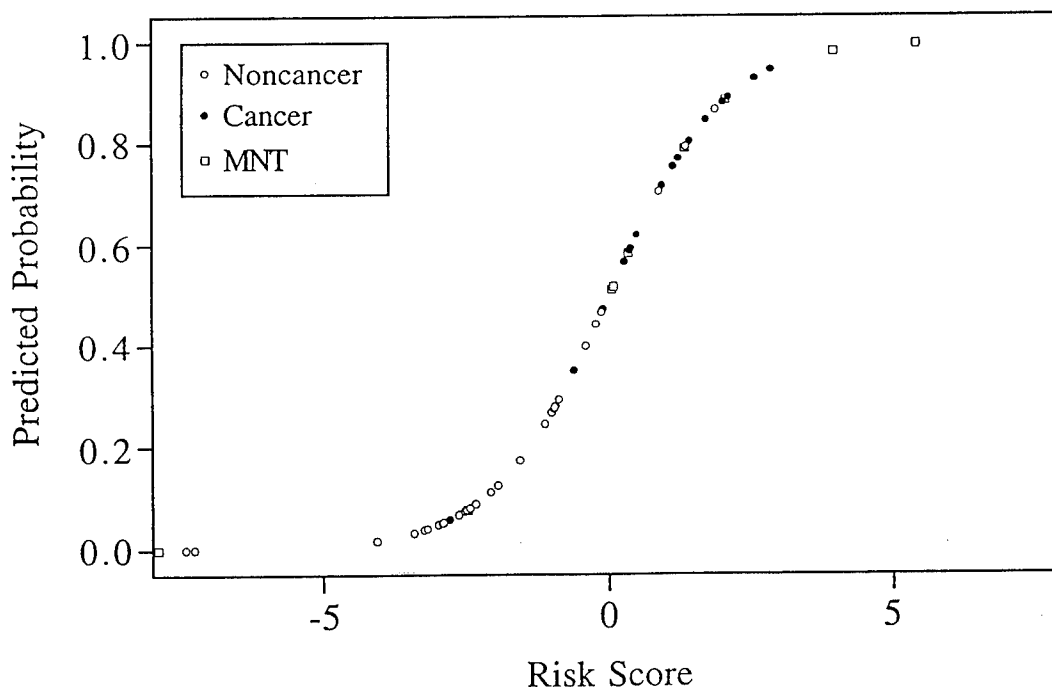


Figure 8. Predicted probability of cancer status using a logistic regression model based on the first six factors from factor analysis.

d. Grouping of patients without cancer

To achieve a spread of groups, the DNA from the 29 reduction mammoplasty patients was grouped based on distance-C (Fig. 9), while blinded to the specific spectral profiles. The three groups were defined by cut-points in distance-C (Group 1 = 0.60 - 0.72; Group 2 = 0.77 - 0.89; and Group 3 = 0.90 - 0.96). The mean values of the \log_{10} ratio of concentrations of Fapy-A/8-OH-Ade were 0.43, 0.73, and 0.087, respectively, which clearly reflects a relative increase in the OH adduct in Group 3. The mean value for the cancer group was -0.31. Thus, changes in the $\log_{10}(\text{Fapy-A/8-OH-Ade})$ ratio among the reduction mammoplasty patients closely reflected changes seen in the FT-IR spectral differences.

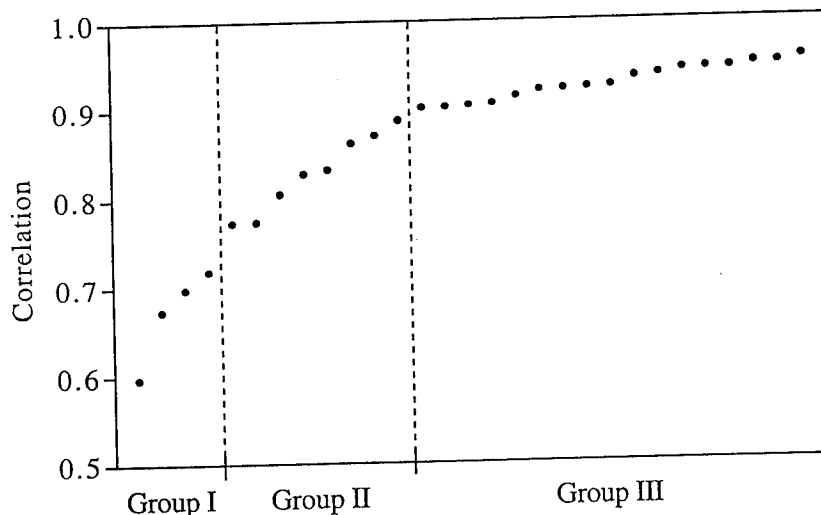


Figure 9. Groups defined by distance from cancer library: far (Group 1, distance-C = 0.60-0.72), intermediate (Group 2, distance-C = 0.77-0.89), and close (Group 3, distance-C = 0.90-0.96), based on distance-C.

Major changes toward a cancer phenotype were evident when comparing groups 1, 2 and 3 relative to the cancer profile (Fig. 10). Group 3, which most closely matched the cancer DNA profile was designated as the cancer-like phenotype and comprised 59% of the total patient database without cancer.

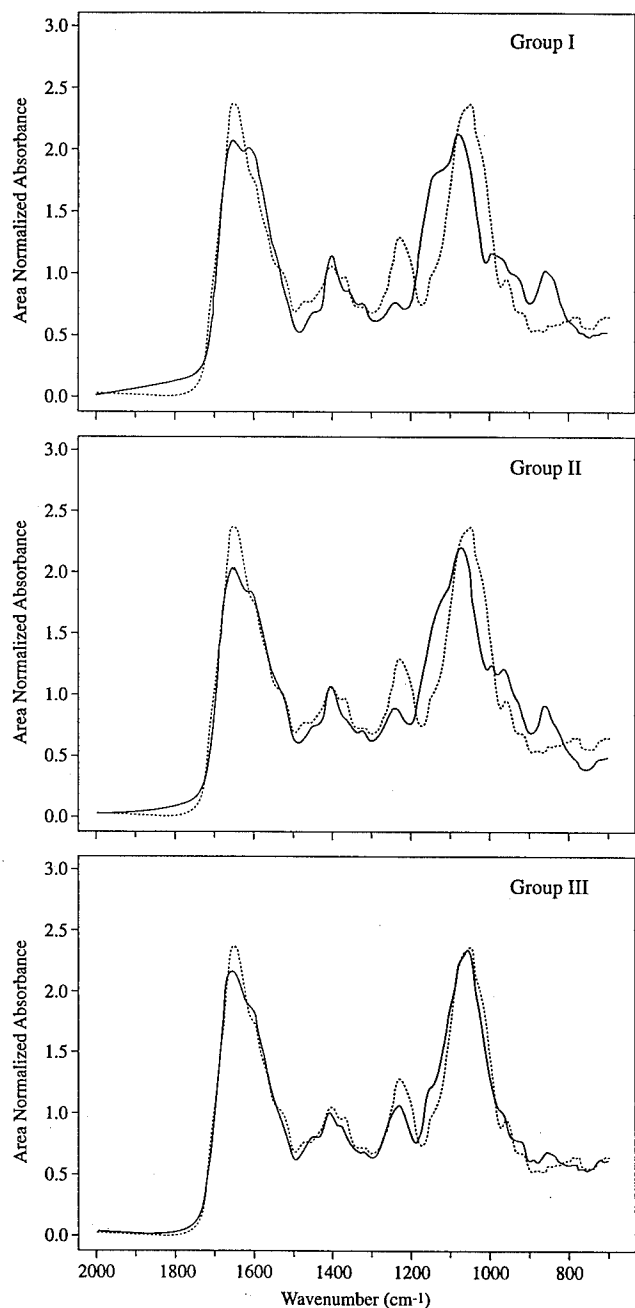


Figure 10. Mean spectra for patients with cancer and for groups of patients without cancer at three distances from the cancer library: far (Group 1, intermediate (Group 2), and close (Group) 3, based on distance-C. The mean cancer spectrum (dashed line) is superimposed on the spectrum of each group.

Beginning with Group 1, progressive absorbance increases were evident in the area 1700-1500 cm^{-1} which is associated with CO stretching and NH_2 bending vibrations in DNA (19, 41). The area between 1550 and 1300 cm^{-1} which is associated with NH vibrations and CH in-plane base deformations showed absorbance increases at $\sim 1450 \text{ cm}^{-1}$. These spectral alterations likely were associated with the $\bullet\text{OH}$ -induced base modifications previously demonstrated in the breast by GC-MS/SIM (9). The band at $\sim 1230 \text{ cm}^{-1}$ which is assigned to the PO_2 antisymmetric stretching vibrations of the phosphodiester backbone (19) progressively developed from a relatively small peak in Group 1 to a relatively large peak in Group 3. As the $\sim 1230 \text{ cm}^{-1}$ peak progressively increased, the shoulder in Group 1 at $\sim 1150 \text{ cm}^{-1}$ became less prominent and showed a slight shift to a lower wave number ($\sim 1110 \text{ cm}^{-1}$) in Group 3. As shown in Fig. 10, the area 1200-1100 cm^{-1} progressively changed between groups until a close match was obtained between the Group 3 spectrum and the cancer spectrum. The area 1200-1100 cm^{-1} was found to have the most statistically significant difference between the cancer and noncancer spectra and in the systematic shift of spectra with distance from the cancer library as shown in the *P* values of Fig. 5 and correlation values of Fig. 6. The ~ 1100 -700 cm^{-1} region exhibited several band shifts and absorbance changes, notably near 875 cm^{-1} , which were assigned to PO_2 - and PO-stretching vibrations of the phosphodiester group and CO-stretching vibrations of the deoxyribose moiety (19). The 875 cm^{-1} area underwent a substantial absorbance decrease from Group 1 to Group 3. An absorbance decrease at $\sim 975 \text{ cm}^{-1}$ was also apparent as the spectral profiles progress toward the cancer phenotype.

e. Results of factor analysis

Standard factor analysis methods were used to determine if the variation across spectra was related to cancer versus noncancer DNA status (Figure 11). In factor analysis, which was used previously in IR spectroscopy (43, 44), the variation among spectra (expressed as sums of squares) is separated into a number of factors, each of which explains some of the variation across the collection of spectra. In the present analysis, the first factor approximately represented the mean spectrum, and other factors represented variations near this approximate mean in decreasing order of importance. Before performing the factor analysis, the decision was made to keep sufficient factors to explain at least 90% of the variation in spectra beyond the variation explained by Factor 1. In these data, the first factor explained 97.1% of the sums of squares of normalized absorbances from 1503 to 761 cm^{-1} . Of the remaining sums of squares, Factors 2-6 explained, respectively, 38.3%, 30.8%, 13.4%, 5.5% and 3.2%.

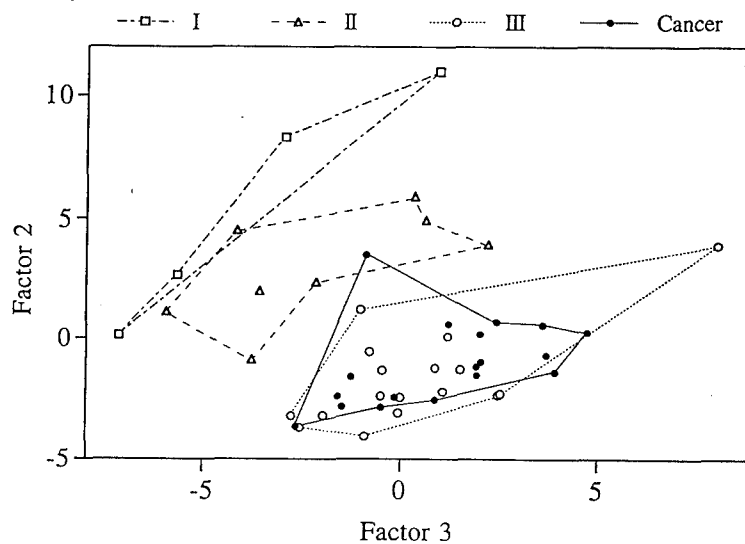


Figure 11. Relationship of cancer and noncancer groups based on factor analysis in the 1503-761- cm^{-1} area. Relative absorbances of the noncancer Groups 1,2, and 3, defined by distance from the cancer library in a separate analysis, overlap little. Group 3 and the cancer group overlap substantially. This plot shows that groups that are coherent relative to distance also are coherent relative to independently derived factors.

Table 7 shows that Factors 1, 2, 3 and 6 differed significantly between patients with and without cancer. The lack of a significant difference between patients with and without cancer relative to Factors 4 and 5 indicates that some of the variation across spectra was not related to cancer/noncancer status. The factors were calculated to explain any variation and, in essence, were blinded to cancer/noncancer status.

Table 7. *Comparison of factor scores between patients with cancer (n=18) and without cancer (n=29) (first six factors).*

Principal Components	Cancer Mean (\pm SD)	Noncancer Mean (\pm SD)	P-value
Factor 1	-30.2 (\pm 0.7)	-29.6 (\pm 1.2)	0.05
Factor 2	-1.0 (\pm 1.8)	0.6 (\pm 3.8)	0.05*
Factor 3	1.1 (\pm 2.2)	-0.8 (\pm 3.1)	0.02
Factor 4	-0.1 (\pm 2.0)	0.1 (\pm 1.9)	0.7
Factor 5	-0.2 (\pm 1.2)	0.1 (\pm 1.2)	0.3
Factor 6	-0.3 (\pm 0.7)	-0.2 (\pm 1.0)	0.05*

SD: Standard deviation.

* *t*-test based on unequal variances.

The cancer and noncancer groups defined earlier were also coherent concerning the factors that were derived disregarding cancer or noncancer status or distance from the cancer library. These factors were derived simply to explain variations among spectra. Fig. 11 shows the groupings in a plot of Factor 2 versus Factor 3, the two most important factors describing variation after Factor 1, which mainly described a mean common to all spectra. There was little overlap among Group 1 (most distant from the cancer library in Fig. 10), Group 2 (intermediate distance) and the combination of Group 3 and the cancer group. Group 3 substantially overlapped the cancer library. Although there were evidently continuous changes defining a progression from noncancer to cancer spectra, two diverse statistical analyses (distance-C and factor analysis) grouped the patients in a similar fashion.

III. CONCLUSIONS

The results describe the ability of GC-MS/SIM and FT-IR spectroscopy to discriminate high vs. low damage in fish liver DNA and cancer versus noncancer status based on differences in intrinsic chemical properties of DNA isolated from female breast tissues. In the case of the breast tissues, the region of the IR spectrum associated with vibrational transitions of structural substituents of nucleotide bases, deoxyribose and phosphodiester moieties defines the most strikingly consistent, nonrandom differences between the cancer and noncancer groups. Thus, the nonrandom variation found most likely was due to inherent chemical properties that are characteristic of each group.

The structural composition and integrity of DNA in living cells and tissues is a matter of prime importance for the fidelity of cell division and the avoidance of mutagenesis. Consequently, it is kept under close scrutiny by repair enzymes (e.g., endonucleases) to eliminate DNA lesions that may result in such changes. In this regard, structural analysis of DNA from a variety of living systems defines several specific types of oxidative chemical aberrations and their frequency of occurrence (6-10; 14). Generally, two types of oxidative modifications are known to occur in DNA, those derived from two-electron oxidations resulting in the formation of generally bulky adducts of the base structures (1-3), and those derived from one-electron oxidations arising from free radical processes resulting in a distinct variety of products, including modifications of base structures and the deoxyribose-phosphodiester backbone (4). The nature and frequency of adducts derived from two-electron reactions of DNA bases have been the subject of intense research for many years (1-3). Such adducts generally occur in the range of 1 in 10^7 to 1 in 10^9 normal bases in a variety of normal and neoplastic tissues (45). Conversely, the products from one-electron oxidations, such as the ring-opening structure Fapy-A, or hydroxylation products, such as 8-OH-Ade or 8-OH-Gua, generally are much more abundant. In the normal breast, for example, the ratio of ring-opening structures to normal bases frequently represents $1:1 \times 10^2$ to $1:1 \times 10^3$ normal bases--a difference of 4 to 7 orders of magnitude relative to the reported values for two-electron oxidation products (3, 45). Thus, the IR spectral differences observed in the present study most probably were controlled substantially by the contribution of one-electron oxidative changes, such as those shown previously to occur in the \bullet OH modification of breast DNA (9). The \bullet OH probably arises from H_2O_2 via the Fe^{++} -

catalyzed Fenton Reaction (4, 5). As previously postulated (9), the H_2O_2 may be formed from the redox cycling of an effector molecule, such as an estrogen derivative or xenobiotic chemicals (e.g., organochlorines) arising from environmental exposure. In this context, the degree of damage may be related, at least partly, to the extent to which normal radical trapping systems in the cell (e.g., reduced glutathione) are overwhelmed by the generation of the $\bullet\text{OH}$ (7, 46).

The spectral profiles obtained in the present study appear to reflect considerable differences in the modification of the nucleotide bases, phosphodiester groups and the deoxyribose moiety between cancer and noncancer groups relative to the attack of the $\bullet\text{OH}$ on DNA. However, other reactions cannot be ruled out. Difficulties presently exist in precisely defining the proportions of damage inflicted on any of these structural components or to specifically identify the exact nature of the functional groups involved. A clearer understanding of these important issues will require detailed study in the future using, for example, oligonucleotides containing known types and amounts of base lesions. Nevertheless, it has been shown that the $\bullet\text{OH}$ attacks the deoxyribose moiety to produce a variety of products resulting from hydrogen abstractions of the pentose ring (47). These reactions result in ring-opening products and the formation of a carbon centered radical at the 5' carbon position, linking deoxyribose residues with the phosphodiester groups. The attack of the $\bullet\text{OH}$ on the deoxyribose moiety is known to result ultimately in strand breakage (47). In this regard, the damage inflicted on the DNA from the normal and cancer tissue was broadly different, as suggested by melting studies conducted in our laboratory which showed that a high proportion of the DNA from the

cancer breast exhibits a series of sub-melts arising from a significant degree of strand breakage. In contrast, the submelts were not evident in the DNA from the normal breast (unpublished data).

It is considerably significant that FT-IR analysis defined a nonrandom progression of changes in the spectral properties of DNA from individuals within the normal group, culminating in a spectrum closely resembling that of the mean spectrum of the cancer DNA. In fact, this progression of spectral changes measured, for example, by distance-C was found to vary directly with measures from a statistical model developed previously (9), based on $\log_{10}(\text{Fapy-A}/8\text{-OH-Ade})$, which compared products from the one-electron oxidation of adenine. The structural alterations measured by IR spectroscopy thus are parallel to the redox-coupled conversions of putatively non-mutagenic Fapy derivatives to mutagenic 8-OH adducts in breast tissues, as defined previously (9). Consequently, each of these analytical methods likely measures a progressive, premalignant condition among women with histologically normal breast tissue as their IR spectra and base lesion ratios approximate those of the DNA from the cancer tissues. We suggest that once such a premalignant condition exists, there is a substantial steady-state concentration of mutagenic DNA lesions maintained by a balance between oxidative attacks and DNA repair processes (9). Then, time-dependent accumulations of mutagenic changes may occur in the DNA. This multistage process ultimately gives rise to conditions that are the necessary and sufficient elements needed for malignant conversion and tumorigenesis. This perspective is consistent with recent results described by Frenkel *et al.* (13) indicating the potential of circulating levels of autoantibodies reactive with 5'-hydroxymethyl-2'-deoxyuridine to be predictive of future incidence of breast cancer among women. These autoantibodies are

presumably elicited by 5'-hydroxymethyl-2'-deoxyuradine derived from an oxidative attack on breast DNA.

In the past 25 years, the risk of female breast cancer has substantially increased to the current 11% lifetime incidence rate. This likely is associated with increased exposures to materials capable of eliciting a mutagenic response and/or dietary deficiencies in protecting against oxidative damage. Given that carcinogenesis is a chronic process driven by an accumulation of random mutational events, that often develop over a number of years, it is expected that a relatively high proportion of the population shows evidence of premalignant changes in the DNA, as found in this report. However, it is expected that only a portion of those individuals, depending on their specific genetic susceptibilities and DNA lesion profiles, would achieve the necessary conditions leading to malignant conversion and tumorigenesis. Therefore, populations residing in areas with variable incidences of breast cancer likely have proportionate differences in the percentage of individuals having premalignant DNA changes of the type described.

Our results predict that future incidences of cancer would arise preferentially among the group of normal individuals whose IR spectra most closely resemble that of the cancer DNA (Group 3; Fig. 10). One way to test this hypothesis would be to conduct a spectral analysis of the DNA from the MNT obtained from nearby the IDC. The noncancer tissue from the ipsilateral breast has been shown to have a high incidence of recurring primary carcinoma (48). A comparison of the DNA from the MNT with the established predictive model showed that 8 of 10 tissues analyzed were classified as cancer. That is, a substantial, distinct group of individuals

known to be at heightened risk for recurrent breast cancer exhibited a pronounced risk on the basis of the DNA model, although the tissues were classified normal on the basis of a routine microscopic examination. Our previous findings (9) showing that the \log_{10} DNA base concentrations and ratios between the IDC and the MNT were not statistically different are completely consistent with these spectral results with the DNA.

The structural modifications in the breast DNA represent a premalignant state almost certainly characterized by various degrees of genetic instability, which was also noted previously from the base-model data (9). This is potentially important in that it has been shown that oxidants, such as the $\bullet\text{OH}$, participate in the activation of protooncogenes and the inactivation of tumor suppressor genes resulting, for example, in an increase in the mutagenesis of hot-spot codons of the human p-53 gene (49). Given the ability of $\bullet\text{OH}$ -induced base lesions to create mutagenic events, it is clear that this process associated intimately with carcinogenesis (9, 14). The demonstrated impact of the $\bullet\text{OH}$ on DNA indicates that familial susceptibility to cancer, defined by cancer-associated genes (50), most probably is influenced by the type and degree of DNA damage demonstrated in the present study.

The extension of the progressive spectral changes observed in the normal breast to overlap those of the cancer group (Fig. 10), which presently is difficult to understand fully, may suggest that cancer risk is likely a continuum relating to the various types and degrees of DNA modification. In this context, the possibility exists that the changes within the cancer group were related significantly to the known constitutive propensity of breast cells to generate high concentrations of H_2O_2 , the apparent precursor of $\bullet\text{OH}$ (51).

A major research emphasis on the observed damage inflicted on DNA seems appropriate in the future, as well as studies directed toward reducing radical concentrations in normal breast tissues so that the DNA repair systems (e.g., the endonucleases) are better able to control or reverse the damage. From the clinical perspective, this may well be accomplished by increasing the intake of antioxidants and the use of drugs containing antioxidant functional groups that preferentially target breast cells. Overall, the findings lend support to the concept that the cellular redox status (9) and $\bullet\text{OH}$ concentrations play a pivotal role in cancer development in the female breast, as previously postulated (9). These conditions may well be reversible and, thus, should be given special consideration in efforts to reduce the incidence of breast cancer.

Aside from providing a new perspective of the etiology of breast cancer, it is apparent that the FT-IR analyses afford a promising means for assessing the DNA status relative to the risk of developing breast cancer. This conclusion is exemplified by using the predictive model for cancer/ noncancer status shown in Table 6 to derive risk scores, such as those given in Fig. 8.

The FT-IR analysis is rapid and requires minimal (μg) amounts of tissue, which are likely obtainable via fine needle biopsy procedures. Results of such an analysis will help in defining groups of individuals of low to high potential risk at the earliest stages of oncogenesis when therapeutic intervention would be especially effective.

In conclusion, 59% of normal women studied from the Puget Sound, Washington area were shown to have a DNA phenotype in the breast representing a premalignant state and thus potentially placing them at high risk for developing breast cancer. In this respect, the degree of cancer risk, which presumably diminishes as the spectral profiles increasingly mismatch those of

the cancer profile, can be calculated on the basis of probability models, such as the one used in this study (Fig. 8).

The proportion of normal women with the cancer-like phenotype is disturbingly high. However, therapeutic approaches that stabilize the redox conditions in breast cells, thus reversing oxidative fluxes, may substantially decrease the cancer risk, as would the successful delivery of sufficient antioxidant or reductant compounds to the breast that would counteract the damaging effects of the $\bullet\text{OH}$. In a broader sense, the GC-MS/SIM and FT-IR spectral models also maybe widely useful in predicting risk for other types of cancer.

Further studies are necessary to determine the degree to which the premalignant state occurs in the breast DNA of women from diverse geographical areas, such as relative to various risk factors [e.g., chemical exposures implicated recently in breast carcinogenesis (52)]. The present findings may at least explain partly the high occurrence rate of breast cancer among women and form the basis for an important new paradigm for the prediction, early intervention and treatment of this disease. Considering the severity of this problem, there is a critical need for a controlled prospective study conducted on the basis of the present and previous (9) findings. This should be directed toward testing the association between the analytical parameters described and various epidemiological factors related to the occurrence of the disease.

IV. FUTURE PLANS

The findings described provide a firm basis for a prospective study testing the hypothesis that the nonrandom progression of oxidative modifications in the normal female breast allow for the formulation of statistical models that are predictive of breast cancer in the population. In such a study, attention should focus on understanding temporal changes in the DNA status of women in relation to breast cancer risk. In addition, the data on the IDC evoke the question of the nature of statistical DNA models derived from other cancer-related breast disease, such as atypical hyperplasia. We believe that future work in this direction should be considered. Overall, the data support the conclusion that the DNA-related changes leading to carcinogenesis are phylogenetically conserved and thus can be exploited in DNA modeling related to a wide variety of hormone- and xenobiotic-induced cancers (e.g., those associated with environmental contaminants).

V. REFERENCES

1. Springer D, Mahlum D, Westerburg K, Hopkins M, Frazier D, Later D, et al.
Carcinogenesis, metabolism and DNA binding studies of complex organic mixtures. In: Cooke M, Dennis AJ, editors. Polynuclear Aromatic Hydrocarbons: Chemistry, Characterization and Carcinogenesis. Columbus: Battelle, 1986:881-92.
2. Langenbach R, Nesnow S, Rice JM, editors. Organ and species specificity in chemical carcinogenesis. New York:Plenum Press, 1983:1-667.
3. Jernstrom B, Martinez M, Meyer DJ, Ketterer B. Glutathione conjugation of the carcinogenic and mutagenic electrophile (\pm)-7 β ,8 α -dihydroxy-9 α ,10 α -oxy-7,8,9,10-tetrahydrobenzo(a)pyrene catalyzed by purified rat liver glutathione transferase. *Carcinogenesis* 1985; 6:85-9.
4. Zhao MJ, Jung L, Tanielian C, Mechin R. Kinetics of the competitive degradation of deoxyribose and other biomolecules by hydroxyl radicals produced by the Fenton reaction. *Free Radic Res* 1994; 20:345-63.
5. Imlay JA, Chin SM, Linn S. Toxic DNA damage by hydrogen peroxide through the Fenton reaction *in vivo* and *in vitro*. *Science* 1988; 240:640-2.
6. Malins DC. Identification of hydroxyl radical-induced lesions in DNA base structure: biomarkers with a putative link to cancer development. *J Toxicol Environ Health* 1993 40:247-261.
7. Malins DC, Gunselman SJ. Fourier transform-infrared spectroscopy and gas chromatography-mass spectrometry reveal a remarkable degree of structural damage in

- the DNA of wild fish exposed to toxic chemicals. *Proc Nat Acad Sci, USA*, 1994; 91:13038-41.
8. Malins DC and Haimanot R. *Aquat Toxicol* 1991b; 20:123-130.
 9. Malins DC, Holmes EH, Polissar NL, Gunselman SG. The etiology of breast cancer: Characteristic alterations in hydroxyl radical-induced DNA base lesions during oncogenesis with a potential for evaluating incidence risk. *Cancer* 1993; 71:3036-3043.
 10. Malins DC, Polissar NL, Nishikida K, Holmes EH, Gardner, HS, Gunselman SJ. The etiology and prediction of breast cancer: Fourier transform-infrared spectroscopy reveals progressive alterations in breast DNA leading to a cancer-like phenotype in a high proportion of normal women. *Cancer* 1995; 75(2):503-17.
 11. Maccubbin AE. in Malins DC, Ostrander GK, editors. *Aquatic Toxicology: Molecular, Biochemical and Cellular Perspectives* Boca Raton: CRC Press, 1994:327-86.
 12. Djuric Z, Simon MS, Luongo DA, LoRusso PM, Martino S. Levels of 5-hydroxymethyl-2'-deoxyuridine in blood DNA as a marker of breast cancer risk. [abstract] *Proc Am Assoc Cancer Res* 1994; 35:286.
 13. Frenkel K, Glassman T, Karkoszka J, Taioli E. Anti-oxidized DNA base autoantibodies as a potential biomarker of high risk for the development of human breast cancer. [abstract] *Proc Am Assoc Cancer Res* 1994; 35:97.
 14. Floyd RA. The role of 8-hydroxyguanine in carcinogenesis. *Carcinogenesis* 1990; 11:1447-50.

15. Kuchino Y, Mori F, Kasai H, Inoue H, Iwai S, Mirua K, et al. Misreading of DNA templates containing 8-hydroxydeoxyguanosine at the modified base and at adjacent bases. *Nature* 1987; 327:77-9.
16. Weitzman SA, Turk PW, Milkowski DH, Kozlowski K. Free radical adducts induce alterations in DNA cytosine methylation. *Proc Natl Acad Sci, USA* 1994; 91:1261-4.
17. Klein JC, Bleeker MJ, Saris CP, Roelen HCPF, Brugghe HF, Van den Elst H, et al.. Repair and replication of plasmids with site-specific 8-oxodG and 8-AAFdG residues in normal and repair-deficient human cells. *Nucleic Acids Res* 1992; 20:4437-43.
18. Koch KS, Fletcher RG, Grond MP, Inyang AI, Lu XP, Brenner DA, et al. Inactivation of plasmid reporter gene expression by one benzo(a)pyrene diol-epoxide DNA adduct in adult rat hepatocytes. *Cancer Res* 1993; 53:2279-86.
19. Parker FS. Chapter 9. In: Applications of infrared, raman, and resonance raman spectroscopy in biochemistry. New York: Plenum, 1983:349-98.
20. Dizdaroglu M, Bergtold DS. Characterization of free radical-induced base damage in DNA at biologically relevant levels. *Anal Biochem* 1990; 156(1):182-8.
21. Franklin RE, Gosling RG. The structure of sodium thymonucleate fibers. I. The influence of water content. *Acta Crystallogr* 1953; 6:673.
22. Jordan DO. The chemistry of nucleic acids. London:Butterworth, 1960.
23. Falk M, Hartman KA, Lord RC. Hydration of deoxyribonucleic acid. II. An infrared study. *J Am Chem Soc* 1963; 85:387-91.

24. Schweder T, Spjøtvoll E. Plots of p-values to evaluate many tests simultaneously. *Biometrika* 1982; 69:493-502.
25. Hosmer DW, Lemeshow S. Applied logistic regression. New York:Wiley, 1989.
26. Troll W, Wiesner R. The role of oxygen radicals as a possible mechanism of tumor promotion. *Ann Rev Pharmacol Toxicol* 1985; 25:509-28.
27. Roy D, Floyd RA, Liehr JG. Elevated 8-hydroxydeoxyguanosine levels in DNA of diethylstilbestrol-treated Syrian hamsters:covalent DNA damage by free radicals generated by redox cycling of diethylstilbestrol. *Cancer Res* 1991;51: 3882-5.
28. Falck Jr F, Ricci Jr A, Wolff MS, Godbold J, Deckers P. Pesticides and polychlorinated biphenyl residues in human breast lipids and their relation to breast cancer. *Arch Environ Health* 1993; 47:143-6.
29. Djuric Z, Heilbrun LK, Reading BA, Boomer A, Valeriote FA, Martino S. Effects of a low-fat diet on levels of oxidative damage to DNA in human peripheral nucleated blood cells. *J Nat Cancer Inst* 1991; 11:766-9.
30. Steenken S. Purine bases, nucleosides and nucleotides: Aqueous solution redox chemistry and transformation of their radical cations e^- and OH adducts. *Chem. Rev.* 1989; 89:503-20.
31. Falck Jr F, Ricci Jr A, Wolff MS, Godbold J, Deckers P. Pesticides and polychlorinated biphenyl residues in human breast lipids and their relation to breast cancer. *Arch Environ Health* 1993; 47:143-6.

32. Breimer LH. Repair of DNA damage induced by reactive oxygen species. *Free Rad Res Comms* 1991; 14:159-71.
33. Tchou J, Kasai H, Shibutani S, Chung MH, Laval J, Grollman AP, et al. 8-oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity. *Proc Natl Acad Sci USA* 1991; 88:4690-4.
34. Cheng KC, Cahill DS, Kasai H, Nishimura S, Loeb LA. *J Biol Chem* 1992; 267:166-72.
35. Pero RW, Roush GC, Markowitz MM, Miller DG. Oxidative stress, DNA repair, and cancer susceptibility. *Cancer Det Prev* 1990; 14:555-61.
36. Robinson E, Rennert G, Rennert HS, Neugut AI. Survival of first and second primary breast cancer. *Cancer* 1993; 71:172-6.
37. Krinsky NI. Effects of carotenoids in cellular and animal systems. *Am J Clin Nutr* 1991; 53:38S-46S.
38. Palozza P, Krinsky NI. Antioxidant effects of carotenoids *in vivo* and *in vitro*: An overview. *Methods Enzymol* 1992; 213:403-20.
39. Weitberg AB, Weitzman SA, Clark EP, Stossel TP. Effects of antioxidants on oxidant-induced sister chromatid exchange formation. *J Clin Invest* 1985; 75:1835-41.
40. Wolf G. Retinoids and carotenoids as inhibitors of carcinogenesis and inducers of cell-cell communication. *Nutr Rev* 1992; 50:270-4.
41. Tsuboi M. Application of infrared spectroscopy to structure studies of nucleic acids. *Appl Spectrosc Rev* 1969; 3:45-90.
42. Miller RG. Simultaneous statistical inference. 2nd ed. 1981:New York: Springer-Verlag.

43. Fredericks PM, Lee JB, Osborn PR, Swinkels DAJ. Materials characterization using factor analysis of FT-IR spectra. Pt 1: Results. *Appl Spectros* 1985; 39(2):303-10.
44. Fredericks PM, Lee JB, Osborn PR, Swinkels DAJ. Materials characterization using factor analysis of FT-IR spectra. Pt 2: Mathematical and statistical considerations. *Appl Spectros* 1985; 39(2):311-6.
45. Randerath E, Randerath K. Postlabeling methods - an historical review. *IARC Sci Publ* 1993; 124:305-14.
46. Ketterer B. Detoxication reactions of glutathione and glutathione transferases. *Xenobiotica* 1986; 16:957-73.
47. Von Sonntag C, Hagen U, Schon-Bopp A, Schulte-Frohlinde D. Radiation-induced strand breaks in DNA: chemical and enzymatic analysis of end groups and mechanistic aspects. *Adv Radiat Biol* 1981; 9:109-42.
48. Fisher B, Bauer M, Margolese R, Poisson R, Pilch Y, Redmond C, et al. Five-year results of a randomized clinical trial comparing total mastectomy and segmental mastectomy with or without radiation in the treatment of breast cancer. *N Engl J Med* 1985; 312:665-73.
49. Hussain SP, Aguilar F, Amstad P, Cerutti P. Oxy-radical induced mutagenesis of hotspot codons 248 and 249 of the human p53 gene. *Oncogene* 1994; 9:2277-81.
50. Hinds PW, Finlay CA, Quartin RS, Baker SJ, Fearon ER, Vogelstein B, et al. Mutant p53 DNA clones from human colon carcinomas cooperate with *ras* in transforming primary

rat cells: A comparison of the "hot spot" mutant phenotypes. *Cell Growth Differen* 1990; 1:571-80.

51. Szatrowski TP, Nathan CF. Production of large amounts of hydrogen peroxide by human tumor cells. *Cancer Res* 1991; 51:794-8.
52. Falck F Jr, Ricci A Jr, Wolff M, Godbold J, Deckers P. Pesticides and polychlorinated biphenyl residues in human breast lipids and their relation to breast cancer. *Arch Environ Health* 1992; 47:143-6.